

## Self-trust and self-confidence: some distinctions

Lloyd J. Carr

Philosophy Department

Rivier University

Nashua, NH 03060

[lcarr@rivier.edu](mailto:lcarr@rivier.edu)

Website: <http://www.rivier.edu/faculty/lcarr>

### 1. Introduction: framework and assumptions

Relations imply relata, and so *interpersonal* relationships require distinct persons, at least two, among whom they hold. One person, for example, might come to the aid of, befriend, joke with, deceive, bargain with, trust, take interest in or respect another person, or one person might have any of these relationships to several other persons who might or might not reciprocate. Similarly, *interagential* relations require distinct agents, at least two, among which they hold; for example, two chess programs might creatively play each other, or a human and Watson might compete at the game *Jeopardy*, or human passengers might trust (or distrust) their plane's automatic pilot, or a family might benefit from being driven somewhere by a self-driving smart-car or living in a smart-house that recognizes and serves each individual inhabitant. For interpersonal and interagential relations, it is generally the nature of the relation that is questioned; it is clear who or what the two or more required relata are.

The same requirement exists for *intrapersonal* relations but in this case the relata, again at least two, are not as obvious as they are in the case of interpersonal or interagential relationships. Self-esteem, self-respect, self-acceptance, self-control, self-doubt, self-deception, self-confidence, self-trust, bargaining with oneself, being one's own worst enemy, and self-denial, for example, are thought to be deeply human possibilities, yet there is no clear agreement about who or what forms the terms between which these relations hold. Philosophers who work in the area of intrapersonal relations have devised two strategies for dividing one person into the required multiplicity such that these are (i) sufficiently diverse for these kinds of self-relations to hold, yet (ii) sufficiently unified to constitute one person – a single self. A problem, of course, is equally satisfying both conditions (but that's another story).

One strategy, Humean in spirit if not by direct influence, is to conceive a person as an aggregate of interests, each a coherent assembly of epistemic and practical abilities and accomplishments; each rich enough to be considered a kind of "sub-self" having its projects, goals, duties and "life." Some of these interests conflict while others are mutually supportive; some function on a base-level while others function on higher-orders; some dominate others and some as it were "voluntarily step aside"; some tend to impulsiveness, some are timid; some take a longer while others a shorter view of issues the person confronts, some are critical of others and some are defensive. But there is no super-interest that

manages, coordinates, or referees all the others. This conception presents a person as more a set of systems each with a degree of independence, rather than a system of sub-systems. A person is left internally “messy,” subject to inner struggles, self-defeating projects and failed attempts at self-management, as well as subject to self-agreements, self-improvement and self-satisfaction. Intrapersonal relationships, on this model, are relations between and among such “parts” of the self. Self-doubt for example has this structure: one of a person’s interests questions or rejects an accomplishment(s) of some other interest of that same person. Take someone, upon driving away from her house for a long vacation, who says to herself, “I remember locking the door, I know I did; but I’d better go back and check just to be doubly sure.” Here, this person’s epistemic interest (seeking assurance) doubts her practical interest’s accomplishment (locking her doors as a responsible homeowner). It is *self-doubt* in so far as the relata are different interests within one and the same person. Donald Davidson, in several influential essays on problems of practical agency, works within this framework.<sup>1</sup> Trudy Govier, in her seminal essay on self-trust, likewise appears to accept such a model of the self.<sup>2</sup> Jon Elster offers a vivid illustration: “I wish that I didn’t wish that I didn’t wish to eat cream cake. I wish to eat cream cake because I like it. I wish that I didn’t like it, because, as a moderately vain person, I think it is more important to remain thin. But I wish I was less vain.”<sup>3</sup> Another illustration: suppose a long-time smoker decides for health reasons to quit smoking. Based on past efforts, he knows that he will have a craving to smoke for a long time after quitting and will end up giving in to the temptation. To manage this un-steadfast interest in quitting, he turns to snacking which does the trick – it overcomes the craving. But now he experiences weight gain that, in his case, has as serious health consequences as his smoking. To manage his weight, he starts an exercise routine; but this takes time, commitment, and dedication if it is going to work. And in the interest of health, his exercise can’t lead to injuries, ... . It is noteworthy that this concept of a person as a population of diverse interests, because of its formal similarity to the world of interpersonal relations, is able to draw on and apply to intrapersonal relations insights from the study of human interactions.

A second strategy, Lockean in spirit if not by direct influence, divides a person into stages within a unit of time. A person is conceived as an aggregate, a composite, of temporally earlier and later selves, each a “snap-shot” of the whole person at a time or at a phase/stage of an epistemic or practical undertaking. Intrapersonal relations are understood as holding between and among earlier and later selves or temporal stages of a person during his or her undertaking. At each stage or “snap shot” a person is a complex assembly of epistemic and practical capacities, interests and accomplishments; stages or selves have access to each other, and each has the ability for reflection, that is: to see another stage of itself both from its vantage point in time, and to see itself from its other stage’s point in time. For example, self-doubt might have the structure of a later self questioning or rejecting the accomplishment of an earlier self of the same person and that earlier self accepting its later self’s judgment. The self-doubter who says to herself, “I remember locking the door, I know I did; but I’d better go back and check just to be doubly sure.” is analyzed as a later self doubting her memory and her certainty of her earlier self’s action of locking her door, and her earlier self agreeing with, not rejecting, this doubt. George Ainslie, Elster, and Thomas Schelling have worked out this model of an agent as a population of time-indexed selves who negotiate a variety of interactions and relationship with each other, in great detail in their different explorations of practical rationality, especially preference instability, self-binding

(precommitment), and self-management.<sup>4</sup> Edward Hinchman's analysis of self-trust and its failures insightfully employs this model.<sup>5</sup>

Each conception of a person or a self as a collection of interacting parts and powers has, of course, a long and varied philosophical history. And each strategy for dealing with intrapersonal relationships is open to (and has received) a variety of criticisms, each resting on several controversial metaphysical assumptions. Granting this, I find the second strategy less problematic and more manageable than the first for my topic and will use it as my framework for distinguishing between self-trust and self-confidence.<sup>6</sup> In what follows, then, I will explore self-confidence and of self-trust, and suggest some distinctions between them, that use temporally earlier and later selves of the same person as relata for these two (easily confused) self-relations.

It will help my efforts to be limited by two restrictions, one pragmatic and the other conceptual. First, to keep my topic manageable, I will focus on these two intrapersonal relations in the context of practical agency and neglect them in the epistemic context (to the degree that these two areas of human effort can be treated separately). Self-confidence and self-trust will here be examined as operating within a practical agent's abilities, rationality and actions, and not examined as operating within an agent's epistemic projects, goals or rationality.

The territory of practical agency is vast and complex. A schematic is needed to locate where self-confidence and self-trust enter the picture. The rational choice view of a practical agent – an ideal rational individual-decision maker who, in pursuit of a goal, calculates the expected utility of her options, each an act-state-outcome trio, and chooses the best option – leaves little room, as far as I can see, for self-trust or self-confidence; so I put this conception aside. For my purposes, then, let a practical agent be composed of the following five capabilities:

1) Cognitive – the ability to form and hold beliefs, including the ability not to believe (withhold belief) about what is or what is not the case; the ability to form and hold expectations, including the ability not to form or to cancel expectations, about what should or should not be the case or what ought or ought not happen. Expectations, then, are normative and are not just anticipations or predictions. I will assume that beliefs can be more or less strongly held, and that: to believe that something is not the case = to disbelieve that it is.

2) Emotive – the ability to desire or not desire that (positively) something be the case or happen, or (negatively) that something not be the case or happen. I take an agent's positive desire to have the power to make the object of desire appear subjectively valuable (that is: desirable) to the agent, and negative desire the power to make its object subjectively dis-valuable (undesirable) to the agent. Having no desire that something happen or that it not happen, then, is not the same as (negatively) desiring it not happen.

3) Volitional – the ability to form an intention, or withhold from doing so, to do or to refrain from doing something. Not intending to refrain from doing an action is not equivalent to intending to do it. I include here the ability to make a decision, and will assume that intentions are formed and decisions made with greater or lesser commitment and resolve.

4) Physical – the ability to act as well as the ability not to act (that is: to hold back and refrain from action) on an intention or on a desire. The “inability” to so act or refrain, then, is not in the sense that an agent physically can’t, but in the sense in which the agent who says, “I can’t” means “I can’t get myself to do it” even though the agent is physically able to do so.

5) Rationality – the ability to bring reason and standards of reasonableness to bear on one’s practical agency, to reflect critically on the exercise of and the products of one’s practical abilities, and a practical agent’s responsiveness to the norms and requirements of rationality with regard to principles of justification, coherence, consistency, and means-ends coordination. This capacity includes the failure of an agent to be rational: to be unreasonable, to have inconsistencies and to be irrational within an undertaking.

This general profile of a practical agent, we see, makes no assumption one way or the other that a practical agent is a morally good person. The next step is to set out a general framework for a unit of practical agency. I will take a practical agent’s undertaking to start at  $t_1$  with an emotive state and end at  $t_3$  with an action, and between the emotive state and the agent’s action comes a volitional state at  $t_2$ :

( $t_1$  emotive state ----->  $t_2$  volitional state ----->  $t_3$  agent’s action).

There are 64 possible combinations. For example: at  $t_1$  Jack desires to lose weight by exercising, at  $t_2$  Jack intends to exercise at  $t_3$ , and at  $t_3$  Jack fails to exercise. Or: at  $t_1$  Jill desires not to go to the dentist at  $t_3$ , at  $t_2$  Jill decides to go to the dentist at  $t_3$ , and at  $t_3$  Jill goes to the dentist. Or: at  $t_1$  Jack wants to watch TV at  $t_3$ , and at  $t_2$  Jack intends not to watch TV at  $t_3$ , but at  $t_3$  Jack ends up watching TV. Or: at  $t_1$  Jill does not desire not to hurt her friend’s feelings, at  $t_2$  she does not form an intention not to hurt her friend’s feelings at  $t_3$ , and at  $t_3$  Jill (says something that) hurts her friend’s feelings. Or: at  $t_1$  Jack desires not to play video games at  $t_3$ , at  $t_2$  Jack intends not to play video games at  $t_3$ , at  $t_3$  Jack does not play video games. Only some of the possible combinations will be used to explore self-confidence and self-trust.

Between an agent’s desire stage at  $t_1$  and the agent’s intention-forming stage at  $t_2$  of an undertaking, additional practical capabilities might, and typically do, operate: beliefs, expectations, other intentions, rationality or other desires. Likewise, between  $t_2$  and  $t_3$  stages of the above unit of practical agency, beliefs, desires, rationality or expectations might, and typically do, operate. This picture of a practical agent requires the assumption of constancy or stability of these capabilities and their contents. So, a  $t_1$  desire to do something is assumed to continue to operate when a  $t_2$  intention to do it is formed, and both still operate at  $t_3$  when the agent performs an action, thereby both discharging the earlier intention and satisfying the earlier desire. Within a unit of practical agency, then, we have a single agent distributed over – “spread-out” into – a series of time-indexed “selves,” each with the full capacity for practical agency; each “self” represents an event of composition – another layer built up or stage added on – in the agent’s undertaking. In this model of a person self-confidence and self-trust are intrapersonal relations to the degree that they are analyzed as relations between and among one agent’s earlier and later selves.

The second restriction is intended to avoid a conceptual (logical) problem of self-reference. If the scope is global, then both self-confidence and self-trust apply to themselves; an agent would be confident about her self-confidence, and be trusting about his self-trust. In the affirmative, such self-reference seems to be conceptually harmless even though practically it might be self-reinforcing to the point that it becomes too difficult for an agent to examine critically and skeptically his excessive self-confidence or to question and escape her excessive degree of self-trust (and why would you, if you are self-confident and self-trusting compounded up several higher orders?)<sup>7</sup> However, if their opposites (as opposed to just their absence) are given the widest scope, self-non-confidence and self-distrust are unsustainable; they are self-undermining. If an agent lacks confidence in his lack of self-confidence, or if an agent distrusts her self-distrust, then the original (base-level) lack of self-confidence or self-distrust are weakened and undermined; they become conceptually as well as practically impossible for an agent to have. Yet it is obvious that agents *do* experience a lack of self-confidence and *are* self-distrusting. To block this problem, I will narrow the scope of self-confidence and self-trust to local domains of action and to specific practical capacities. An agent will be described as self-confident or not, self-non-confident or not, about a given action or course of action that the agent performs or intends to perform, but not described as self-confident or not about everything within the agent's realm of possibility. Likewise, an agent will be described as self-trusting or not, self-distrusting or not, about certain of her abilities and projects, but will not be described as self-trusting or not *tout-court*.

Give these two restrictions within the theory of a person as an aggregate of temporally relativized selves, I believe self-confidence operates between an agent's intention and action, and will be seen to run primarily from an earlier self to a later self with respect to an action and only derivatively from a later self back to an earlier self (see note 16). In contrast, I will argue that self-trust operates primarily between an agent's desire and intention-formation, even though self-trust carries through to an action. Also, self-trust, in contrast to self-confidence, will be seen to involve a core of abilities that continue from an earlier to a later self and as such are closely connected to the person's identity through time; a self-trusting practical agent is a specially constituted person, a self-confident agent isn't.

As a final introductory remark, I note that the philosophical literature on self-trust – to the degree I'm aware of it – generally treats this intrapersonal relation as something good and self-distrust as a bad thing. I suspect the same about self-confidence and its lack, especially within the psychological literature, though I'm less familiar with the literature on this topic.<sup>8</sup> I believe this is one-sided; self-trust might be a good thing in some circumstances, but it can also exist in, and tend to strengthen the resolve of, the murderer, the terrorist, and the tyrant. In such cases, self-distrust is far better morally than self-trust. And I'm sure that a Hitler or a Stalin had accurate self-confidence, not just over self-confidence, in many of their horrible undertakings; too bad, we might wish, they didn't have much less of it, better none at all. In distinguishing between self-trust and self-confidence, I will try to separate for each form from content; it is only with respect to their content, I will argue, that we can judge self-trust and self-confidence a morally good or bad relationship a person can have with him-or-herself.

## 2. The problem

Why distinguish between self-trust and self-confidence? In every-day language these two are used interchangeably in many contexts, apparently without any harmful results. One might even argue that conceptually they must be close, for we often think of self-doubt as the equally opposite attitude to both self-confidence and self-trust. Also, it is difficult to separate them conceptually; examples of cases in which a person has one and not the other don't readily come to mind. Doesn't this indicate that they imply each other, even if they are not simply two ways of saying the same thing? Philosophers who work in the area of self-trust have argued powerfully and insightfully for the importance of this self-relation to a number of features and capabilities they consider central to our humanity. Govier, for example, argues that self-trust is a necessary condition for autonomy and self-respect, and a sufficient condition for self-esteem.<sup>9</sup> Keith Lehrer likewise argues that self-trust is foundational for both epistemic rationality and practical wisdom.<sup>10</sup> Hinchman considers self-trust the basis of our exercise of practical reason.<sup>11</sup> Given that these philosophers are correct that self-trust is something central to human agency, to the exercise of our autonomy, to our epistemic and practical sense of self, and perhaps to a person's overall sense of self-worth, it is essential that it not be confused with other intrapersonal relations; even if some are closely related to self-trust, they might not have the core importance that it is argued to have. Yet when following the arguments and analyses of these philosophers, one gets the feeling that they are sometimes referring to self-confidence not to self-trust, or are using "self-trust" and "self-confidence" interchangeably as if they were two expressions for the same thing. Here, for example, is Govier:

Nor on the present account is self-trust to be valued absolutely. A person may in several different senses have too much self-trust. One may have more confidence in one's own character and capacities than past experience and evidence would warrant, thus having "too much" self-trust in an epistemic sense.<sup>12</sup>

Hinchman, for another example, considers two situations: in one is an agent who is not confident about her memory of an earlier action, and in the other the agent is not confident about an earlier decision; for Hinchman, if these cases of lack of confidence in an earlier self's action or decision are reasonable, then the agent's mistrust in her earlier self's action and of her earlier self's decision are reasonable. But, Hinchman argues, if the lack of confidence in her earlier self's action or decision are unreasonable, the mistrust in each case is "pathological."<sup>13</sup>

In these representative passages, I am not questioning the claims or the arguments being made; I only mean them to show that here we have insightful and influential work on self-trust – argued to be at the very core of practical agency – in which the target concept of self-trust is sometimes linked to, or interchanged with, or "blended" with that of self-confidence as if there were no problem in doing so, no argument needed. Contrary to this impression, I believe that self-trust and self-confidence can and should be distinguished; that even though they might have structural and dynamic similarities (making them subject to being confused) there are nevertheless important differences between them as to their role in and value for practical agency.

### 3. Self-confidence: a proposed analysis

What exactly does an agent who lacks self-confidence lack self-confidence about (or in)? The answer “everything” has been ruled out as unsustainable, if not impossible. It must be something specific to the agent in question, not someone else, for it would then be a case of lack of confidence (in someone else) and not a case of lack of self-confidence. Let’s consider some typical examples.

-- a business manager must make a presentation to his staff and his superiors, and worries he is not fully up on the topic and will do poorly

-- a teenager plans to lie to her parents about her use of drugs; as she thinks up the perfect story, one she is sure her parents will accept, she starts to worry that she won’t be able to pull off her elaborate deception in a convincing enough manner

-- a competitor in an athletic contest has just witnessed a stunning performance, and starts to feel that when it’s her turn she will do badly and won’t “measure-up”

-- a recent college graduate is about to be interviewed for a desired job, and is afraid he will make a poor impression

-- a student in a communication class must make a class presentation, and experiences stage-fright at the thought of doing poorly in front of her classmates and her teacher

-- an admirer wants to ask that attractive and popular person for a date, but loses nerve at the thought of being turned down

-- a new driver plans to go on a long car trip, but feels too inexperienced and worries about highway-speed driving on unfamiliar roads and gives up her plan

-- an elderly person has been given a new computer and believes she’ll be confused by this technology, and is afraid that she’ll do something wrong and damage it.

-- a homeowner has purchased an item of furniture that requires assembly; told the assembly would be easy, he now finds the instructions confusing and the task daunting and fears he will fail to put it together correctly

It seems natural to describe these examples as in each case a lack of self-confidence, but it seems odd to call them instances of self-distrust or even failures of self-trust. Given that these count as typical cases of lack of self-confidence (or self-non-confidence), what is their common form? In each we have:

(i) an agent who intends to, or has committed to, or is required to perform an action. Prior to forming the intention the agent might or might not have a desire to do the intended action; having a prior desire is, thus, not necessary and will be left out of my analysis.

(ii) the agent believes that the action in question is “public” in the sense that there are social, perhaps official, norms/standards for judging how well or poorly the action is performed, that it is not solely up to the agent who is to do the action to determine how well or poorly or correctly

it is done,

(iii) the agent desires to do well: that is, to have the action judged “successful,” or “good,” or “acceptable” or “correct” by the applicable norms, and

(iv<sub>a</sub>) the agent believes (for various reasons) he will fail (or will be judged to have failed) to perform as the agent desires, that is, well by the applicable norms.

I take these four conditions to be the core of a lack of self-confidence, even though a variety of emotions and physical symptoms – worry, fear, stage-fright, nervousness, panic, loss of nerve, avoidance, sweat, stomach in knots – would typically accompany an agent’s belief that she will “publically” fail to do well at something she desires to do well, and believes might be desired or expected by others who matter to the agent that she do well. (We will see below that self-distrust is not accompanied by such symptoms.)

Given that conditions (i)-(iii) hold, self-confidence requires a change in condition (iv<sub>a</sub>) to:

(iv<sub>b</sub>) the agent believes (for various reasons) that s/he will perform the intended action as the agent desires to perform it; that is, to perform well and be judged a success by the applicable norms.

There are two variables here: the strength of the agent’s belief that s/he will do well, and how well the agent believes s/he will perform the intended action as judged by the applicable norms. Which do we consider self-confidence: an agent who weakly believes s/he will perform exceptionally well, or an agent who strongly believes s/he will perform minimally well? I think it is intuitively clear that the latter has more self-confidence than the former. Thus, the stronger an agent holds a belief that his intended action, if done, will be (judged) good, even if this belief is false, the more self-confident the agent will be. The crucial (iv<sub>a</sub>) and (iv<sub>b</sub>) beliefs, I note, are self-predictions that need not be justified by evidence; self-confidence or its lack need not satisfy epistemic norms of justifications, much less truth, to operate within practical agency. So, on this analysis it is possible for an agent to be self-non-confident even when by her own reflection she finds that there is no (good) reason for her to be. I now examine more closely and argue for each of these four conditions on self-confidence.

Condition (i) seems necessary; if an agent has no intention of doing anything, then it is hard to see how the agent could have (or lack) self-confidence about anything. And if an agent decides not to do a given action (A), and to do something else (B) instead, it is equally hard to see how the agent could have any confidence in herself doing A well (or doing A at all, for that matter). Thus, it follows that if an agent has self-confidence, then the agent intends to perform an action (even though the agent may change her mind at a later point and revoke her intention, perhaps out of a loss of self-confidence). Intending to do an action, of course, implies the belief that the action intended is physically possible for the intending agent to do; I will assume that the intending agent has formed this belief.

The connection in this context between an agent’s intention to perform an action and the action the agent ends up doing, assuming it is performed, isn’t clear. For one, the action an agent intends to do must be in the agent’s future, it can’t in the agent’s past, for it must be possible that the action an agent intends to do never gets done. (It will have to be seen in what sense an agent could have self-



confidence about a past action the agent performed. See note 16.) Thus, we have in condition (i) an agent who at one time intends or decides that at a future time this same agent will perform an action. I see no reason why the intended action must be completely voluntary (that is, internally motivated); the agent might conceivably be under some external obligation, duty, or even coercion to perform it (say, if the agent legally contracted to do it and will now be subject to legal coercion to perform as contracted).

Another point to note is that the action an agent intends to do and the action the agent performs should match<sup>14</sup>; if not, the action couldn't carry out that earlier intention, though it might carry out some other earlier intention. In "matching," the earlier intention to do it becomes at the later time when it is performed, so to speak, the "identifying or guiding principle" or set of constraints that keeps the action being done identical (within bounds) to the action that was intended – if type, then to type; if token, then to token. If the action performed is to execute the earlier intention to perform it, then the intended action must be done *as* it was intended, if not also *because* it was intended (the "because" case, for example, might be an agent who does something merely because they intended to do it; say, merely to prove a point in answer to someone who doubts this agent's intentions are "for real").

Also, if this match or link didn't happen, the action performed couldn't serve as the measure by which the agent's earlier degree of self-confidence was too-much, too little, or accurate. Consider this case: suppose an agent (S) at  $t_1$  (i.e., the  $t_2$  point in the basic unit of agency in part I above) forms an intention to deliver a speech to an audience at  $t_3$  and is confident at  $t_1$  that her performance at  $t_3$  will be a huge success, and suppose that at  $t_2$  S has a mini-stroke that causes her completely to forget her  $t_1$  intention, and as it happens S ends up for an entirely different reason delivering a speech to an audience at  $t_3$  but it is not the speech S at  $t_1$  intended to deliver. What connection does this agent's  $t_3$  self, the one giving the speech, have to this agent's  $t_1$  confident self? These are an earlier and a later stage of the same agent, but there seems to be no connection, or at least not the right connection, between S's confidence at  $t_1$  in her future  $t_3$  self and S's performance at  $t_3$ , because there is no longer the right match between the action that is intended at  $t_1$  and the action that is performed at  $t_3$ . Even if at  $t_3$  S delivered the exact same speech word-for-word, to the same audience to which S intended to talk at  $t_1$  (say the mini-stroke destroys all short-term awareness S has of her  $t_1$  intention, but not her long-term memory of that speech given countless time to countless audiences); at  $t_3$  S's action would be the one intended at  $t_1$ , but it could not be "identified" as such by the agent applying her earlier  $t_1$  intention, and so performed "under the guidance of" that intention. What connection, in this case, exists between S's  $t_1$  confidence in her  $t_3$  self and her  $t_3$  performance? Again, because the link between S's  $t_1$  intention and her  $t_3$  action has been disturbed, if not completely broken by the stroke, S's  $t_3$  performance can't retroactively confirm or falsify this agent's  $t_1$  level of confidence in her  $t_3$  self; the performance would be as if the earlier intention never existed – it would not be *that* action and the constancy requirement assumed above has been broken for her intention.

Condition (ii) seems necessary; if an intended action is not subject to norms or standards of performance, then there is no way the action can fail; it couldn't fail even if it were not done, for there would be no norm even of completion. And if an intended action is not subject to norms, then it can't be judged or evaluated as to how well or poorly or correctly it is performed; it might comparatively be judged, say, to have been completed faster or slower than another action, but couldn't comparatively

be better or worse unless speed was normative. If an agent believes an action cannot succeed or fail and cannot be normatively evaluated, even if not done, then there is nothing about the action about which an agent who intends to do it could be confident or lack confidence. It follows that if an agent has self-confidence, then the agent believes the intended action is subject to evaluation by norms of performance.

The question that condition (ii) raises is: why must the norms or standards of performance be “public”? Why can’t an agent’s intention be (or contain) the only norms by which an action succeeds or fails to execute that intention, or be judged to have? Why condition self-confidence (or the lack of it) on anything beyond an agent’s will or on anything outside an agent’s control? The answer, I believe, has something to do with fear and the difficulty understanding self-fear. If there is nothing about performing an action that can make an agent worry or be concerned about, if an agent has “nothing to fear” in doing something, then it would seem that any standards of success and failure – any norms of performance – would have to be completely private, subjective, determined and controlled solely by that agent: that is to say, easily manipulated by the agent. Given that there is no such thing as fear without an agent-relative uncontrollable power or force of some kind, it follows that there can’t be self-fear without an inner force or power, some part of the self, the agent can’t control. Now, an agent’s intentions can’t be such an uncontrollable power or force without making agency (at least rational agency) itself impossible. It would seem to follow, then, that self-confidence, as well as the lack of it, requires norms of performance that are in some sense “external” to the agent; that is, “public” in the sense of not completely determined or controlled by the self-confident agent. So, for instance, an agent answerable only to self-created and self-imposed norms of performance could, and would be foolish not to, simply change the norms to fit the performance, thereby making every intention successfully executed. Self-confidence and self-non-confidence become meaningless in this case; they would not be “felt” by the agent as “operating.”

Condition (iii) is necessary; if an agent did not desire to act in a way that satisfies the applicable norms of performance, if an agent who intends to perform an action doesn’t care about and has no interest in its success or failure, then there is no basis or motive within the agent for worry or concern about failure, and no basis within the agent for performing according to the norms in question. Condition (iii) brings the norms of performance, as it were, into the agent’s circle of interests, motivations, and goals. Thus, a self-confident agent is an agent who does not only believe that there is such a thing as doing well in an intended action, and not only predicts that she will do well, she also wants to do well even if prior to forming the intention the agent does not desire (or desires not) to do the intended act. But an agent who wants to do well need not be a self-confident agent. Thus, condition (iii) is necessary for self-confidence. Desire of course implies awareness of what is desired; if an agent who intends to do something desires that it be done well, the agent must “know” – have some idea of – what a successful doing in that case is, or believe that someone able to evaluate its success or failure knows what the standards are. That is to say, the agent must either be aware of the norms of performance in order to desire to act in a way that satisfies them, or believe that there are such norms of which someone is aware who will be able to evaluate the agent’s performance.

Condition (iv<sub>b</sub>) is necessary for self-confidence; it seems impossible for an agent to intend to perform an action and at the same time believe s/he will never do (or try to do) the intended action. To believe the intended action will at some point be attempted, if not completed, by the agent seems part of what it means for an agent to intend to do an action. Thus, if an agent satisfies condition (i), then the agent believes s/he will at least try to do the action the agent intends to do. And, if an agent satisfies condition (iii), the agent desires to do that intended action well. But the content of the agent's desire [*I am judged to do the action I intend to do well*] and the content of the agent's belief [*I will do the action I intend to do*] can't generate self-confidence unless these contents match. Given that degrees of confidence attach to beliefs and not to desires, the agent's belief content must match the agent's desire content, not the other way around. Thus, it follows that if an agent has self-confidence, then the agent believes that s/he will do the intended action as s/he desires to do it – that is, judged a good or correct performance by the applicable norm (and this is condition (iv<sub>b</sub>)).

Given that an agent satisfies conditions (i)-(iii), condition (iv<sub>a</sub>) is necessary for lacking self-confidence. The argument parallels that for condition (iv<sub>b</sub>). In brief: the agent's basic belief content [*I will perform the action I intend to perform*] must conflict with – can't match with – the agent's desire content [*I am judged to do the action I intend to do well*]. The belief content, then, must be [*I will fail to perform the action I intend to perform as I desire to perform it*]. And the agent, we see, can't in this case change her lack of self-confidence into self-confidence simply by making her belief fit her desire, because the agent believes that the norms of performance by which her action will be judged are "out of her hands."

By way of clarification, condition (iv<sub>a</sub>) harbors a potential confusion. We would not accept an agent who says, "I intend to pay back the loan and I desire to pay back the loan, but I do not believe I will pay back the loan"; this doesn't make sense. But we would accept an agent who says, "I intend to pay back the loan and I desire to pay back the loan, but I believe I will not (be able to) pay back the loan." We wouldn't lend this honest agent money precisely because what the agent says makes sense. Thus, condition (iv<sub>a</sub>) for a lack of self-confidence says: the agent believes s/he will *not* perform the intended action well and will do poorly by the applicable norms of performance. It does not say: the agent does *not* believe s/he will perform the intended action well, for a state of non-belief cannot provide a place for a lack of self-confidence (or, for that matter, self-confidence) to operate.

We might ask at this point: doesn't condition (iv<sub>a</sub>) clash with condition (i)? Why would an agent who lacks self-confidence ever even form an intention to perform an action in the first place? Doesn't lacking self-confidence operate prior to intending, and doesn't it block forming an intention to do something the agent would be self-non-confident about doing? Clearly, this can't be the case. As argued here, self-non-confidence is not a belief an agent has about everything the agent might do, it is a belief an agent has about himself relative to a specific performance and its "public" evaluation. Often, the agent has no choice about the action at issue, it is something the agent is required to do by the agent's circumstances or it is something the agent has already made a commitment (perhaps legally binding) to do. Where this is not the case, the agent's self-non-confident belief must still be about something; minimally the agent must, prior to the operation of any lack of self-confidence, consider or present to himself the action at issue as something "I might intend to do." The agent in this case, as it were, imaginatively form the intention to do it in order to be able to discover if this is a performance about

which the agent would lack self-confidence. Without some such prior “intention,” there couldn’t be anything about which the agent would have a self-non-confidence belief.

An added problem for an agent under condition (iv<sub>a</sub>) is that in the case of lack of self-confidence the agent who can escape from acting will probably try to do so. An agent who lacks self-confidence, then, might believe that she will be more severely judged for lacking self-confidence, or for opting out, than for performing the action poorly. This creates a dilemma that the self-confident agent does not face.

I note that both conditions (iv<sub>a, b</sub>) do not require that the intended action is actually performed by the agent; self-confidence and its opposite are beliefs, by this analysis, not properties of actions. But in cases in which the intended action never gets done, the self-confidence (or its lack) does not operate counterfactually. By the above analysis, the person who never swims yet claims, “If I were to swim, I would be (judged) a very good swimmer.” counts as a boastful agent, perhaps, or a bragger, but cannot count as a self-confident agent; likewise, the agent who never swims and claims: “If I were to swim, I would be (judged) a terrible swimmer.” might be mistaken or might be correct in attributing this possibility to himself, but couldn’t qualify as lacking self-confidence about swimming well. In contrast, we will see below that self-trust and self-distrust supports counterfactuals.

How, then, do conditions (i)-(iv<sub>b</sub>) fit together to form an account of self-confidence? I have argued that each condition is separately necessary for self-confidence, that (i)-(iii) are each necessary for conditions (iv<sub>a, b</sub>), and that (i)-(ii) are necessary for condition (iii). Schematically, we have the following analysis of self-confidence for agent (S) and action (A): S has self-confidence about A’ing iff:

- (i) S intends to A,
- (ii) S believes A’ing will be evaluated by “public” norms and expectations of performance,
- (iii) S desires to A in a way that satisfies its norms and expectations
- (iv<sub>b</sub>) S believes S will, in A’ing, satisfy S’s desire.

S has a lack of self-confidence (has self-non-confidence) iff S satisfies conditions (i)-(iii), and:

- (iv<sub>a</sub>) S believes S will fail to satisfy S’s desire (by doing A poorly or by opting out if possible).

This analysis still leaves it unclear (at least to me) why an agent who meets all 4 conditions has *self*-confidence as opposed to just confidence that others will judge his/her A’ing in a certain way, or confidence that A will be done well by an agent with whom s/he identifies. To see more clearly how this concept of self-confidence (I leave out self-non-confidence) might develop – be built up layer-by-layer in a unit of agency – as an intrapersonal relation, I apply the model of a person as a series of time-relative selves, each possessing the capacity for practical agency. For person (S), we have S’s t<sub>1</sub>, t<sub>2</sub>, t<sub>3</sub>, ... t<sub>n</sub> selves. For any given t<sub>n</sub> self, higher number selves represent S’s later selves and lower number selves represent S’s earlier selves within a given unit of practical agency. Person S is, under this conception, a composite of selves developing over a given length of time within a unit of agency, unified by the access each self has as a practical agent to S’s other selves and the influence each self has as a practical agent over the practical capabilities of S’s other selves. For person (S) and action (A) the analysis above (with added commentary) becomes dynamic:

(i)  $S$ 's  $t_1$  self forms the intention that  $S$ 's  $t_2$  self perform action  $A$ . And,  $S$ 's  $t_2$  self agrees to accept and to be bound by  $S$ 's  $t_1$  self's intention that  $S$ 's  $t_2$  self do  $A$ . (That is: at  $t_1$   $S$ 's intention serves as a directive, perhaps even a command or a "telling," to  $S$ 's  $t_2$  self that commits or binds  $S$ 's  $t_2$  self to doing  $A$ .<sup>15</sup> And  $S$ 's intention is consistent and coherent; when  $S$  intends to do  $A$ ,  $S$  believes: that future agent, me, who is to do  $A$  consents to do that action and will not when the time comes for no good reason refuse to do it; and  $S$  believes this because  $S$  believes that the agent who forms the intention and the agent who carries it out by doing  $A$  are the same practical agent,  $S$ , who is not agential self-destructive or self-defeating, inconsistent or incoherent.)

(ii)  $S$ 's  $t_1$  self believes that  $S$ 's  $t_2$  self's  $A$ 'ing is perceived by others, and that when or after  $S$ 's  $t_2$  self performs  $A$ , they will evaluate – judge –  $S$ 's  $t_2$  self's  $A$ 'ing by norms and expectations of performance that neither  $S$ 's  $t_1$  self nor  $S$ 's  $t_2$  self can manipulate at will.<sup>16</sup>

(iii)  $S$ 's  $t_1$  self desires that  $S$ 's  $t_2$  self performs  $A$  in a way that is evaluated positively and not judged negatively. (That is: in desiring that one's future actions meet the norms and expectations that are to be applied,  $S$  is "saying" that it matters to  $S$  that  $S$ 's future performance will be perceived and be judged, and it is of value to  $S$  that it be judged a good or a correct performance.)

(iv<sub>b</sub>)  $S$ 's  $t_1$  self believes (with variable strength) that  $S$ 's  $t_2$  self performs  $A$  as least as well as  $S$ 's  $t_1$  self desires  $S$ 's  $t_2$  self to perform  $A$ . (That is: this belief depends on the above three conditions being fulfilled, and it is this special belief that carries the confidence (of variable strength) in one's future self that makes it self-confidence. It is natural to think of self-confidence, in this model, as varying with the strength of  $S$ 's  $t_1$  condition (iv<sub>b</sub>) belief: the stronger this belief, the more self-confident the agent.)

The above analysis of self-confidence and lack of self-confidence has these beliefs operating in a unit of practical agency between the points of forming an intention to do an action (point  $t_2$  in the general structure of practical agency described in part I above) and the agent's attempt to perform the action ( $t_3$  in this general structure, assuming the intended action is attempted). I would like to note several features about this analysis.

First, it is formal in the sense that  $S$  and  $A$  are variables: " $S$ " ranges over practical agents as its values and " $A$ " over actions that practical agents are able to perform. In its formal structure and dynamics, we see that self-confidence is morally neutral; neither a good nor a bad thing. As stipulated above, it allows well-intentioned as well as evil-intentioned persons to have self-confidence or to lack it. Any given instance of self-confidence or its lack, however, has content; who  $S$  is, what  $A$ 'ing is, and what the practical context or situation is, are specified. Instances of self-confidence or its lack, in this case, are open to moral evaluation either by the agent himself or by any other interested party. So, for example, the terrorist's, the murderer's, the tyrant's self-confidence are, I will assume, morally reprehensible, and their lack of self-confidence a source of hope that these agents might back down. The surgeon's, the trial lawyer's, the student's self-confidence can be judged morally praiseworthy or not, and their lack of self-confidence perhaps a cause for worry. Being "open to moral evaluation," of course, does not mean that such instances are morally evaluated; there is no part or moment in the structure or the dynamics

of self-confidence or its lack (that I can see) that represents, necessitates, or motivates its moral evaluation. It might even be the case that strong self-confidence, and all-the-more over-self-confidence, makes moral reflection both more important and more difficult for the agent to undertake.

Another point of interest is that according to this analysis, self-confidence is not essential to practical agency. It is a self-relation that depends on the capacity for intrapersonal relations, a capacity no doubt essential to our humanity, but it itself appears to be a contingent as opposed to an essential property of agency. The absence of self-confidence (as opposed to self-non-confidence) is clearly not destructive or essentially limiting to practical agency and its presence does not provide an essential link between an agent's intention and the agent's action that carries out that intention. It is possible that the pleasure and assurance that typically accompanies self-confidence, and the trouble, added weight and sense of insecurity of action that typically accompanies self-non-confidence, can help or hurt the agent in performing the intended action, but are not necessary to perform or fail to perform it well; it is easy to imagine the self-non-confident agent's being judged to have done well and the self-confident agent's being judged to have done poorly. And, it would seem, in much of meaningful human life the actions of rational practical agents are done successfully without self-confidence and without self-non-confidence; done perhaps out of habit, routine, automatically and with minimal conscious forethought. And on average we seem to get by. So then, where might the value of an agent's self-confidence or lack of it be found, if not within practical agency itself? Perhaps their value is more social than intrapersonal; to the degree that interagential (including interpersonal) relations require actions and interactions that are (a) governed by norms of performance and (b) performed well, there seems to be an important role for self-confidence, and its lack can be significant.

I note, as a third point of interest, that this analysis allows for the possibility for both self-confidence and lack of self-confidence to become self-sustaining. Take an agent who lacks self-confidence and who performs A well. The fact that the norms of performance are in some sense "public" might go a long way toward overturning S's false negative prediction about how well she A's. S, perhaps after several such well-done A's, could gain increasing self-confidence about A'ing. But now, as typically happens, suppose that disturbing emotions accompany S's lack of self-confidence – S worries about A'ing, becomes nervous, or experiences anxiety, stage-fright or panic, becomes unfocused and easily distracted, and perhaps physically stressed – such that these states interfere with S's doing A well. The probability could well increase that S does A poorly for situations in which S has no choice about not A'ing, and that S fails to do A at all for situations in which the original intention to do A can be overridden by a later decision. And doing poorly would tend to confirm and perhaps even increase S's lack of self-confidence. Self-non-confidence can become, as it were, a self-fulfilling prediction, a self-reinforcing self-relation, to the detriment of the person, as practical agent, who lives with it (but, I should add, may be a benefit to the agent's society for cases in which A is morally condemnable). The same self-sustaining loop is possible for self-confidence. The accompanying emotions typical of self-confidence – feeling secure, calm, focused, on-top of the situation and in control, feeling mentally sharp, perhaps being physically sharp – could tend to enhance S's performance and increase the probability of a favorable evaluation. This, in turn, would confirm and perhaps even increase S's self-confidence in the given domain of action (which, in the case of the terrorist, for example, is to be regretted, feared and condemned).

Finally, this analysis provides a way to understand the overly self-confident agent, the agent whose self-confidence is too-little, and the difference between having too little self-confidence and having a lack of self-confidence (self-non-confidence). We see that too-much self-confidence is not a problem within condition (iii): the agent's desire to do well by the applicable norms of evaluation; rather it is a mismatch between conditions (iv<sub>b</sub>) – the agent's belief that the agent's future self will A in a way that satisfies condition (iii) – and how well the agent actually A's or the evaluation that future self receives for A'ing. Being overly self-confident is either to perform less well than the agent's former self believes the agent's future self will perform, or to believe more strongly that the agent's future self will A well than is warranted (say, it was a stroke of luck that S does A that well, not to be repeated). The greater this mismatch is, the more overly self-confident the agent. Under-self-confidence, then, is mismatch in the other direction: the agent's future self performs better, by the norms of performance the agent accepts, than the agent's earlier self believed his future self would perform. Under-self-confidence, on this understanding, is quite different from a lack of self-confidence. An agent who has self-non-confidence satisfies condition (iv<sub>a</sub>); but an agent who has too little self-confidence satisfies condition (iv<sub>b</sub>), only in a way that mismatches her future self's performance evaluation.

#### 4. Self-trust: a proposed analysis

The concept of a practical agent as a set of time-indexed selves among which intrapersonal relations hold allows, as was mentioned above, insights, mechanisms and techniques from the world of interpersonal relations to be used in the area of self-relations. So, as people can bargain and negotiate with each other, we can bargain and negotiate with ourselves; as we have ways of justifying our beliefs or actions to others, we can use these to justify our beliefs or action to ourselves; and, as people have certain techniques and means by which they seek to manage or govern others, a person might be understood to use similar techniques and means in seeking self-management or self-government.<sup>17</sup> Philosophers who have worked on the problem of trust as an interpersonal relation have used their insights, analyses and models to understand self-trust. Here, for example, is Govier:

Usually we think of trust in interpersonal terms, as an attitude one person has toward another, or as an important aspect of a relationship between people. To show that self-trust is a kind of trust, I seek to explain central features of interpersonal trust and then show that these features are also present in an attitude toward oneself that can properly be called "self-trust."<sup>18</sup>

Hinchman, likewise, models self-trust on trust in others based on their analogy:

It is revealing that self-trust runs the same risk of pathology as trust in others. In each case, I'll argue, reasonable trust presupposes trustworthiness in the trusted. But reasonable mistrust equally presupposes untrustworthiness in the mistrusted, or evidence thereof. That's why the exercise of practical reason is incompatible with baselessly failing to trust your earlier intention forming self. Yet earlier selves are not inherently more trustworthy than other people. If it is ever reasonable to act on the basis of self-trust, it must sometimes be reasonable to act on the basis of trust in others.<sup>19</sup>

I accept this methodology and in what follows try to apply certain features of interpersonal trust to the case of self-trust.<sup>20</sup> However, there is an important feature in one's relation to oneself that isn't found, and cannot exist, in one's relationships to other people: a special kind of intimacy, degree of transparency and expectation. In the area of self-relations, it is hard to shake the "Cartesian insight." When all is said and done, one person can't "know" another person as well or as deeply or as closely as a person can "know" herself (even though in both cases there might well be more confusion than clarity, more ignorance than insight); one person's expectations about another person do not have the same kind of force as one's self-expectations, and thus the realization of frustrated or violated expectations in each case are not equally disturbing or disturbing in the same way; one person's rightful powers over another person are not as sweeping as one's rightful powers over one's self, and so self-resistance and self-rebellion and failures of self-control are not the same as interpersonal resistance and rebellion; a person for whom another is a problem is in a very different position than a person who is a problem to themselves; it is one thing to dislike or to be fed-up with or disgusted with or ashamed of someone else, but it a very different thing to dislike or to be fed-up with or disgusted with or ashamed of one's self. These vaguely stated truisms are meant to underscore the deep disanalogy between inter- and intrapersonal relations.<sup>21</sup> Applying results from the study of trust in other agents to the study of self-trust, therefore, can only go so far. At some point, the exploration of self-trust must part ways with that of interagential trust if it is going to reveal the place trust has in a practical agent's relation to his-or-herself.

There are three points of disanalogy that I believe are important. First, in interpersonal trust (but not in interagential trust, for example when human agents trust autonomous smart-machine agents such as the fully self-driving car), it is always possible for the one trusted to betray the trust placed in them. The betrayal of trust is a different type of trust-risk in trusting other persons from the risk that the trusted turns out to be untrustworthy and one's trust is disappointed; betrayal is a violation of trust, not just a let-down of trust. In contrast, self-betrayal does not seem possible; a self-trusting agent may be untrustworthy and let herself down, but the "Cartesian insight" does not seem to allow in self-trust for the opacity, lack of inner access, secrecy, and perhaps malice, between different persons that are required for betrayal. The impossibility, then, of self-betrayal in the case of self-trust and the possibility of betrayal in the case of trusting other (human) agents make these two kinds of trust significantly different. In this respect, self-trust is closer to interagential trust than it is to interpersonal trust.

The second dissimilarity between intra- and interagential trust (including interpersonal trust) centers on control and autonomy. Trust in other persons and non-person agents implies that the trusted agent is autonomous, at least in the domain of activity in which s/he or it is trusted; if one agent completely controls the actions of another agent such that the latter has no autonomy (in the sense of the ability to weigh options and act on its own, without being controlled by other agents) there is no room for the one to trust the other to do an action or for the other to function as trustee. However, an agent can't have self-autonomy without making agency impossible; an agent (whether a human or a smart-machine) requires a degree of self-control or self-regulation, s/he or it can't be "beyond control" or "out of control" with respect to the scope of its own powers or s/he couldn't trust themselves – and it couldn't be trusted – to do anything. The impossibility of self-autonomy in the case of self-trust and the



necessity of trustee autonomy in the case of trusting other agents make for an important disanalogy between these two kinds of trust.

The third significant disanalogy involves the kinds of thing interagential (including interpersonal) trust and intrapersonal trust are. I believe (and have argued: see Carr 2012) that trusting other agents, when it is rational to do so, are individual decisions based on justified beliefs. Trusting or distrusting another agent is a choice an agent makes, an act of good or bad judgment founded on beliefs about the other agent but not itself another belief. For example, when you enroll in a retirement fund that periodically re-balances your investment profile automatically as you age, you trust this autonomous e-agent to make financially responsible re-adjustments of your retirement savings within the bounds of financial risks that apply. You don't just believe this about the fund and the algorithms controlling it, for you can believe this without enrolling in the fund; the decision to enroll is a decision to trust. Justified trust in another agent is a good decision, and unjustified trust a bad decision. In contrast to this, in what category do we intuitively (the argument comes below) place self-trust? It seems wrong to say that you decide to trust or to distrust yourself; self-trust and self-distrust are more in the nature of beliefs you have about yourself than they are choices you make about yourself. When it comes to one's self, the interdependence between trusting yourself and belief in your own trustworthiness is more intimate than in the case of trusting other agents; the psychic space that separates them in interagential trust, making room for the needed decision, collapses in intrapersonal trust. Trusting one's self, when there is good reason to do so, is not evaluated as an "a good choice" based on belief, it is itself believing something about one's self (or an enduring self-attitude for ongoing self-trust). If this is correct, then – as a decision and a belief are importantly different – there is an important third disanalogy between interagential trust and self-trust.

When an agent is self-distrusting, what is it about the agent that is distrusted? "Everything" has been ruled out as an unsustainable, if not conceptually impossible; and the target of distrust can't be something about someone else, for then the distrust wouldn't qualify as a self-relation. Consider the following examples:

-- an elderly person explains his new hearing aid by saying, "I didn't want to admit it, but I could no longer trust my hearing, I reached the point where I was misunderstanding everyone."

-- a student is taking a final exam on topics she feels weak in, and is sitting next to someone who has mastered the material; during the exam she thinks: I know I shouldn't peek, but I can't trust myself not to glance over at his exam to see what his answers are; I can't resist looking.

-- a person who works around money has gotten into debt and thinks: I know I shouldn't, but it would be so easy that I can't trust myself not to take some of this money when no one is looking; I can't keep from taking some, I really need it.

-- a person who can't stand the sight of blood and feels faint in its presence is asked at a traffic accident to help stop an injured person's bleeding; he thinks: I want to help, but I can't trust myself not to get sick or maybe faint if I do; I just can't help it.

-- late at night on a long drive home alone, a tired driver thinks: I want to make it home for a good night's sleep, but I don't trust myself to stay awake; I'm not able to fight my drowsiness.

-- a suicide bomber, as the time of her act of martyrdom draws near, starts to feel she lacks the courage and conviction to go through with her mission, and secretly worries: can I trust myself to go through with this?

These are examples for which self-distrust seems the right description; it would be odd and unnatural to describe them as instances of a lack of self-confidence. If we add in the typical and sensible mechanisms and strategies people use to cope with their self-distrust, many more examples come to mind:<sup>22</sup>

-- the late-sleeper doesn't trust himself to get out of bed for work in the morning and not be late again, so he moves the alarm clock away from the bed and makes it as loud as it will go

-- the cigarette addicted mother doesn't trust herself to keep her promise to her children to stop smoking, so she flushes an expensive stock of cigarettes down the toilet

-- the overweight man doesn't trust himself to keep to his exercise plan, so he joins a health club that charges a stiff fee for failing to show up for sessions in the program in which he enrolled.

-- the suicide bomber doesn't trust himself to follow through with his mission, so he boosts his courage with another exposure to the propaganda of his cause and takes drugs as the time approaches.

Such strategies, it seems to me, do not make up for a lack of self-confidence, and they don't provide the agent with self-confidence; they are better understood as ways for an agent to substitute external mechanisms for internal areas of self-distrust. Granted, then, that these are all acceptable examples of self-distrust, consider their common form. In each we see:

(i) an agent who desires to perform or refrain from performing an action "A" (e.g. hear others, not copy answers, not take money, stay awake, not get sick, exercise, commit an act of terrorism, etc.);

(ii) the agent believes that satisfying this desire (i.e. doing or refraining from doing the action) is in some way challenging, difficult or minimally not easy for the agent; the agent believes it requires the agent to be capable: to have an ability, skill, resolve or to make an effort, or to have a certain strength of ability, skill, effort or resolve, and that this capability is not just enabling with respect to A'ing or refraining from A'ing, it is also productive of A'ing or refraining from A'ing;

(iii) the agent believes s/he should possess the requisite capability: the (strength of) ability, skill, effort or resolve to satisfy condition (i) desire; the agent values possessing this (strength of) ability, skill, effort or resolve and normatively expects it of him or herself;

(iv<sub>a</sub>) the agent believes s/he does not possess the capability in question; the agent believes s/he is not sufficiently capable to meet his or her own expectations; as a result,

(v<sub>a</sub>) the agent does not form an intention to satisfy condition (i) desire, or intends not to satisfy it, or – if possible – turns to an external mechanism that substitutes for the missing capability so that condition (i) desire is satisfied “with assistance” (e.g. a hearing aid, a loud alarm clock, destroying opportunity to back out, threats of a fine, drugs, etc.)

I will take these 5 conditions to capture the core form of self-distrust. The distrust, we see, comes primarily with conditions (iv<sub>a</sub>), a consequence of which yields condition (v<sub>a</sub>). To capture self-trust, then, we must change (iv<sub>a</sub>) to:

(iv<sub>b</sub>): the agent believes s/he possesses the capability: the (strength of) ability, skill, effort or resolve, and consequently change (v<sub>a</sub>) to:

(v<sub>b</sub>): the agent forms an intention to satisfy condition (i) desire, with no external assistance needed. I now argue for each of these 5 conditions and examine each more closely.

Condition (i) is necessary; if an agent did not desire to do an action (A) or refrain from doing an action (not-A), then there would be no internal reason (no motive internal to the agent) to consider whether or not the agent was up to the task of A'ing or not-A'ing (though there might be a theoretical reason).<sup>23</sup> But self-trust and self-distrust, by analogy with trusting and distrusting other human or non-human agents, intuitively involve accessing an agent's ability to A or to not-A; an agent not able to A can't be trusted to A. Thus, if an agent is self-trusting (or self-distrusting), then the agent desires to perform or refrain from performing an action.

Why, it may be asked, say that the agent *desires* to A or *desires* to refrain from A'ing? Why not state condition (i) as the agent intending or deciding to A or refrain from A'ing? One reason is the logical coherence of the 5 conditions: unlike desire, intending or deciding to A implies belief in the ability to A; this conflicts with condition (iv<sub>a</sub>) for the case of self-distrust, and duplicates condition (v<sub>b</sub>) for the case of self-trust; the set of 5 conditions wouldn't be coherent. The other reason is phenomenological: self-trust and self-distrust seem to be experienced as based on a belief about the ability or lack of ability to do something an agent wants to do – and would like to be able to do – rather than on what an agent might intend or decide to do without any accompanying desire to do it, or perhaps with a desire not to do it.<sup>24</sup>

Condition (ii) is necessary; we would have to believe that “desiring = its own satisfaction” – to believe that “to want is to get” – in order to consider condition (ii) unnecessary. Clearly, this is only the case for a human agent who desires to desire: the desire is automatically self-satisfied and the issue of trust or distrust doesn't arise. Otherwise, this seems true: if (a) an agent who desires to A is the agent who satisfies that desire by doing A, then (b) that agent must possess any ability, skill, resolve or strength of effort necessary to do A or refrain from A'ing and at least one that is sufficient to do A or refrain from A'ing. Now if an agent satisfies condition (i), then the agent satisfies the antecedent (a) of this conditional; we have an agent who desires that s/he, not someone else, satisfy her desire by doing A or by refraining from A'ing. It follows this agent satisfies the consequent (b); that agent must have the

capability to A or refrain from A'ing. But why would this agent make this into a belief? What would internally motivate forming such a belief? If, for example, A'ing or refraining from A'ing takes courage or alertness or perseverance, then why would an agent who desired to A or refrain from A'ing form the belief that s/he must possess that courage or alertness, or perseverance? Why would such a belief be formed by the same agent who satisfies condition (i)? The reason, I believe, is that the agent considers A'ing or refraining from A'ing in some sense challenging or difficult for the agent. Self-trust and self-distrust wouldn't operate in a unit of practical agency that an agent found easy to do or refrain from doing, something the agent ordinarily does. Any agent who believes a task would be difficult for that agent to do, and who also desires to do it, would also form the above consequent (b) as a belief; and this is an agent who satisfies condition (ii). Thus, condition (ii) is necessary for self-trust and for self-distrust.

Condition (iii) is necessary; by condition (iii) a self-trusting and a self-distrusting agent evaluates him-or-herself with respect to a capability; namely, the capacity the agent believes (by condition (ii)) is required to A or to refrain from A'ing. Importantly, the agent does not evaluate his or her present or future behavior; it is a judgment about capability, not performance.<sup>25</sup> If the agent did *not* believe that the (strength of) ability, skill, effort or resolve (hereafter shortened to "ability" or "capability") needed to satisfy the condition (i) desire *should* be part of his-or-her powers of practical agency, it would be difficult to see how the condition (i) desire could be anything more than an empty wish. All the more: if the agent believed that the required ability *should not* be part of her powers of practical agency, this belief would clash with and undermine practical agency by rendering the condition (i) desire moot. We would not readily accept an agent who says, "I desire to do A, and believe x ability is needed to do A, and I believe I should not have this ability." This makes no sense, unless we could charitably interpret this agent as desiring to do something he wishes he not be able to do, perhaps an agent trying to keep himself from doing a (believed) wrong he uncontrollably desires to do.<sup>26</sup> Thus, it follows that if an agent is either self-trusting or self-distrusting, then the agent believes s/he should be capable of satisfying his or her condition (i) desire.

We can see two dimensions in this self-evaluation: one concerns the ability the agent believes is required and productive to satisfy condition (i) desire and the other concerns the agent's normative self-expectation with respect to possessing this ability. With regard to the first, the agent clearly values the capability, and believes it is something good. In the context of practical agency, this value is minimally instrumental (whatever additional value the ability might have for the agent) and subjective; the capability in question, then, is good in the agent's eyes as a means to do or refrain from doing A – that is: valuable as a means to satisfy the desire of condition (i).<sup>27</sup>

With regard to the second, an agent who believes that she *should* have a given capability to do or refrain from doing an action imposes or applies a normative expectation on herself as a practical agent. She, as it were, holds herself up to a certain standard of practical agency with regard to doing or refraining from doing the action. The agent judges herself, we might say, from the vantage point of an ideal that she holds, and this ideal is normative for this agent to the degree that she applies it to herself; that is, the agent agrees with this ideal's "command" that she ought to be the kind of practical agent who has that ability.<sup>28</sup> By what right or on what grounds, we might ask, does the self-trusting or the self-distrusting agent impose such a normative self-expectation? What if the self-expectation is unrealistic? Clearly,

desires alone can't justify self-expectations. Just because an agent desires, say, to run a marathon does not justify this agent expecting he should have the ability to do so. The answer, I believe, is that the basis for a normative self-expectation lies in the agent's autonomy and is constrained by the agent's practical rationality. A reasonable practical agent will have realistic self-expectations; unrealistic self-expectations, in turn, indicate an "imperfectly rational" practical agent.<sup>29</sup> An autonomous practical agent will, by definition we might say, have the right to self-govern. A non-self-governing agent can't be an autonomous agent. Thus, if an agent is self-trusting or self-distrusting, then the agent satisfies condition (iii); and any agent who satisfies condition (iii) is – or at least takes him-or-herself to be – with respect to the capability required to satisfy condition (i) desire, a sufficiently autonomous and reasonable practical agent.

Condition (iv<sub>a</sub>) is necessary for self-distrust; this condition says that the agent finds an inner weakness in his practical agency, a sort of psychic incapacity or "moral" disability. This is not an inability to perform an action due to "external" causes; the agent sees himself deficient in something the agent reasonably expects he should possess, something obligatory by his own standards. A practical agent who believes that he is not in fact capable in a way that he believes a practical agent in his situation ought to be, (1) can't – and shouldn't – trust himself to be a practical agent who can exercise that ability, and (2) must distrust himself to act on the basis of that capability. This agent, in other words, believes – i.e., judges or evaluates – himself to be untrustworthy in a domain of activity. This belief is special in two important ways. First, it is general: an agent who satisfies conditions (i)-(iv<sub>a</sub>) must consider himself not worthy of *anyone's* trust when it comes to exercising a capability he believes he does not possess (strongly enough). Second, it is counterfactual: an agent who is self-distrusting in this way believes that he would be equally untrustworthy in *every* similar unit of practical agency were they to arise (but in fact don't), for he believes he still can't trust himself to do what he desires to do even when – because of his self-distrust – he intends not to do it and in fact doesn't do it; self-distrust lasts through an action that doesn't get performed, it does not cease with a decision not to A or not to refrain from A'ing. Thus, a self-distrusting agent is one who believes himself untrustworthy in a given *domain* of exercising an ability and associated range of *possible* action. An agent, then, who is self-distrusting clearly satisfies condition (iv<sub>a</sub>), but an agent who satisfies condition (iv<sub>a</sub>) need not be a self-distrusting agent, for this agent might not satisfy any of the other conditions on self-distrust. Thus, condition (iv<sub>a</sub>) is necessary for self-distrust.

Why would an agent who satisfied conditions (i)-(iii) also satisfy (iv<sub>a</sub>)? Why wouldn't such an agent just refuse to form or accept a belief about such a perceived inner deficiency? This is of course a possibility, but I believe unlikely. For one, it is doubtful that will or decision or even desire governs belief; doubtful, that is, that practical reason can dominate epistemic reason "at will." But I grant that this is controversial. However, given that the agent is (or takes herself to be) rational (see condition (iii)), the agent will have evidence that s/he is deficient in the ability in question. If we return to the representative examples at the beginning of this section on self-trust – the elderly man, the student, the sleepy driver, the late-sleeper, the overweight person, etc. – we see that they are practical agents whose beliefs in their own lack of ability or lack of inner strength are typically epistemically justified by a degree of self-knowledge based on, say, a history of experiences in which the incapacity or lack of inner

strength were made manifest. In this case, self-distrust is justified by the agent's accurate belief in her own untrustworthiness. Refusing to believe in one's inner weakness given such evidence would make for a very unusual, confused and troubled practical agent, and if this agent were to intentionally (falsely) presents himself to others as not internally "disabled" in this way, he thereby acknowledges his belief that he is agentially deficient.

A bigger problem, I believe, is that this self-belief might be inaccurate, it might "miss the mark" and an agent could distrust herself to a lesser degree (or possibly greater degree) than she is untrustworthy. I take this to be self-*mistrust*: the possibility of being overly self-distrusting or too-little self-distrusting; the agent's degree of self-distrust does not match the evidence the agent has of her untrustworthiness; equivalently, the agent is more or less untrustworthy than she distrusts herself to be with respect to the capability at issue. So, for example, an agent who believes he lacks the strength to stay awake when driving at night – who believes he won't be able to fight off his sleepiness – and who believes he ought as a practical agent, undertaking a night-driving task, have the ability to keep awake and successfully fight off such drowsiness, considers himself untrustworthy to continue driving and arriving home, which is what he desires. But this agent might get it wrong; he might be able to stay awake with strength of effort he does not realize he has, based on a history of too-quickly giving in to his sleepiness. This agent is self-mistrusting in the sense that his self-distrust is excessive; he is somewhat less untrustworthy than he believes himself to be with respect to the ability to stay awake on his late night drive home. (But, we recognize, this might be a fortunate case of being wrong in a good way, erring on the side of caution, a case of inaccurate self-distrust that qualifies as "better safe than sorry.")

Condition (iv<sub>b</sub>) is necessary for self-trust; condition (iv<sub>b</sub>) stands to self-trust as condition (iv<sub>a</sub>) stands to self-distrust. (iv<sub>b</sub>) says in effect that the agent believes herself sufficiently capable in an agential quality that the agent believes she ought (as a practical agent in that given task) to possess. A practical agent who believes that she does in fact possess the ability she believes a practical agent in her situation ought to possess, (1) can – and should – trust herself to be a practical agent who *can* exercise that capability, and (2) can and should trust herself to act according to that capability (given that the agent also believe that circumstances have not changed so as to create unanticipated problems). This agent, then, believes – judges or evaluates – herself to be trustworthy. We can see that this belief is both general and counterfactual. An agent who satisfies conditions (i)-(iv<sub>b</sub>) must consider herself worthy of *anyone's* trust when it comes to exercising that ability she believes she possesses, and is self-trusting in the range of *possible* action that would depend on that capability were any to be desired and intended. It follows that a self-trusting agent is one who believes herself trustworthy in a given *domain* of exercising an ability and range of associated *possible* action. A self-trusting agent, then, satisfies condition (iv<sub>b</sub>); but an agent who satisfies condition (iv<sub>b</sub>) need not be a self-trusting agent, for example if this agent did not satisfy any of the other conditions on self-trust. Thus, condition (iv<sub>b</sub>) is necessary for self-trust

Because self-trust is typically taken to be a positive part of practical agency (by the agent herself as well as others who might benefit from her actions), there is not the worry that arose in the case of self-distrust that a practical agent might not want to admit she is self-trusting, and resist this belief; there is instead the opposite worry: a practical agent might be disposed to holding a false belief that she is trustworthy in the exercise of an ability when she isn't. Thus, there is the rational requirement that the

agent justifies her self-trust and critically reflects on its accuracy; that is, it should be supported by evidence that the agent is trustworthy, and such evidence would presumably come from the agent's self-knowledge based on a history of experience exercising the ability at issue. As with self-distrust, self-trust might "miss the mark." It is possible for an agent to be self-mistrusting, that is: overly self-trusting or too-little self-trusting, because there is the possibility that the degree of self-trust does not match the agent's degree of trustworthiness with respect to exercising that ability. The chain-smoking mother, for example, who doesn't destroy her stock of cigarettes because she (overly) trusts herself to keep her promise to her children to quit, perhaps believing that her love for them and her promise to them are much stronger than her smoking habit (and in her experience have always proved to be much stronger than most things in her life), may regret she didn't destroyed them when her resolve approaches the breaking point in its clash with her addiction, and her children discover her sucking on an unlit cigarette.

Given that conditions (i)-(iv<sub>a</sub>) have been satisfied, the agent is self-distrusting; a reasonable consequence is condition (v<sub>a</sub>). Similarly, given that conditions (i)-(iv<sub>b</sub>) are satisfied, the agent is self-trusting and condition (v<sub>b</sub>) would result. It is reasonable to conclude that a self-distrusting reasonable agent will form the intention not to exercise a capability she believes herself untrustworthy to exercise. Thus, the sleepy driver does not continue driving but pulls over for a rest; the elderly person whose hearing is going gets a hearing aid and does not continue to struggle making out what folks are saying. Likewise, the self-trusting reasonable agent will form the intention to satisfy her condition (i) desire by A'ing or refraining from A'ing, but of course might fail to do so for various external or internal reasons.

Of the above 5 conditions on self-trust, then (i), (ii), (iii), and (iv<sub>b</sub>) have been argued to be necessary for self-trust and (v<sub>b</sub>) argued to be the result. And for self-distrust: (i)-(iv<sub>a</sub>) are each necessary conditions and (v<sub>a</sub>) would result, or so I've argued. In sum, where "S" ranges over practical agents, "A" ranges over actions, and "X" ranges over an agential capability (an ability or strength of ability, skill or strength of skill, effort or strength of effort, resolve or strength of commitment) S is self-trusting with respect to X and A iff:

- (i) S desires to A or to refrain from A'ing,
- (ii) S believes that A'ing or refraining from A'ing requires X,
- (iii) S normatively expects, as a reasonable autonomous practical agent, that S should possess X,
- (iv<sub>b</sub>) S believes that S does possess X (and as such is trustworthy with regard to A'ing or refraining from A'ing); consequently,
- (v<sub>b</sub>) S forms an intention or makes a decision to A or to refrain from A'ing.

S is self-distrusting with respect to X and action A iff S satisfies conditions (i)-(iii) and:

- (iv<sub>a</sub>) S believes that S does not possess X (and as such is untrustworthy); consequently,
- (v<sub>a</sub>) S does not form an intention to A or refrain from A'ing, and might form an intention not to A or not refrain from A'ing, or might when possible intend to substitute "external" assistance for S's inability/incapacity.

Given that these 5 conditions get at the structural core of self-trust, I would like to turn to the dynamics of self-trust as an intrapersonal relation. How does self-trust get built up, what are its layers, and function in a practical agent's temporally forward and backward coordination of its multiple time indexed selves, and function to unify these selves such as to form a relationship one has with one's self? Using the above analysis as a base and adding comments, I offer the following: for practical agent (S), agential capability (X), and action (A), at times  $t_1, t_2, t_3, \dots, t_n$ , S is self-trusting with regard to X and to A'ing or refraining from A'ing iff:

(i) S's  $t_1$  self desires to A or to refrain from A'ing, such that S's  $t_1$  self desires that S's  $t_3$  self desires to A or to refrain from A'ing, and S's  $t_1$  self desires that S's  $t_3$  self, not anyone else, satisfy S's  $t_1$  self's and S's  $t_3$  self's desire by A'ing or by refraining from A'ing at  $t_3$ . (That is: it is not simply that S desires to do something, it is that S desires that S satisfies S's desire, not some other practical agent, and that S's later self still desires what its former self desires. The late sleeper, for example, who desires at night to get up in the morning in time for work, desires at night that (a) in the morning it is she, not someone else, who gets herself up in time for work, and that (b) in the morning she still desires to get up in time for work and not that in the morning she no longer desires to and instead desires that it is her partner, not she, who wants her to get up in time for work (even if this is true). Condition (i) establishes a practical agent's desire that forms the base for self-trust and that has stability over the unit of time during which practical agency is exercised.)

(ii) S's  $t_1$  self believes that S's  $t_3$  self must have X for S's  $t_3$  self to satisfy S's  $t_1$  self's desire, and S's  $t_3$  self accepts this belief. (That is: S's later self who attempts to A or to refrain from A'ing does not reject as false S's earlier belief that S's later self requires capability X in order to A or refrain from A'ing, but shares that belief. There is constancy to the agent's belief about X from its earlier to its later self and back.)

(iii) S's  $t_1$  self normatively expects that, as a practical agent, S's  $t_3$  self have X, and S's  $t_3$  self believes that S's  $t_1$  self is both reasonable and autonomous in normatively expecting this of S's  $t_3$  self. (That is: S's earlier self expects that S's later self *ought to be capable* or *ought to be sufficiently disposed*, as a functioning practical agent of a certain kind, to A or to refrain from A'ing; and S's later self believes that its earlier self has every right – is completely justified – in imposing on and obligating it with this expectation; from the viewpoint of S's later self, S's earlier self is within its rights to tell it how things should be with it.)

(iv<sub>b</sub>) S's  $t_1$  self believes that S's  $t_3$  self in fact possess X and thus meets its expectations, and S's  $t_3$  self believes that S's  $t_1$  belief is correct. (That is: S's earlier self believes with good reason that S's later self is (sufficiently) trustworthy with respect to A'ing or refraining from A'ing, and so trusts S's later self to do so. And S's later self believes with good reason that its earlier self is (sufficiently) trustworthy with respect to the normative expectation it places on it and with which it binds it, and so trusts its earlier self in having done so.) Given these conditions,

(v<sub>b</sub>) S's  $t_2$  self commits or enlists S's  $t_3$  self to do or refrain from doing A by forming the intention that S's  $t_3$  self do or refrain from doing A, and S's  $t_3$  self by accepting this commitment attempts to A or refrain



from A'ing within the constraints and guidance of S's  $t_2$  self's intention. (That is: an earlier and a later self of the same agent, in holding reciprocal justified beliefs in each other's trustworthiness on the matter of A'ing (though not with respect to the same thing), have a relation of mutual trust holding between them. As stages of the same practical agent's undertaking, this qualifies as self-trust. From the earlier to the later self, it is trusting the agent's ability and determination – i.e., capability – to A or to refrain from A'ing; from the later back to the earlier self, it is trusting the agent's intension that it satisfy a desire. It is as if in forming the intention one says to the other, "Don't worry, you are in good hands, I will not get you into trouble." and the other says back to the one, "And don't you worry, you are in good hands, I will not let you down." except that it is one person "talking" to himself, so it means: "I trust myself, and I am trustworthy, on the matter of A'ing or refraining from A'ing." This  $t_2$  intention, then, is founded on the agent's self-trust)

Self-distrust and degrees of self-mistrust can likewise be understood as dynamic intrapersonal relations by putting conditions (i)-(v) into this same framework: that of a practical agent as a composite of coordinated time-indexed selves, interrelated by their access to, their ability to see themselves from the other's perspective, and their effect on each other's capabilities for practical agency. However, I leave this step undone and, before moving to a discussion of some distinctions and relations between self-trust and self-confidence, turn to the following noteworthy points about self-trust.

1) Self-trust, on this analysis, is at the center of the exercise of a certain important form of practical agency, but is not essential to all forms of practical agency. It seems possible for an agent to act with no prior desire or even intention to so act, and by "automatically" and "thoughtlessly" acting produce a result such that this qualifies as a unit of agency (perhaps the agent can in some sense even be held legally responsible for that result) – no self-trust operating or needed. However, given that prior to acting there is desire, belief, and intention, which form the typical unit of practical agency, self-trust has been argued here to play an essential part of successful agency.

2) On the level of form, self-trust as integral to practical agency is not inherently morally good or bad, as practical agency itself is formally neither necessarily morally good nor bad. The moral dimension, we see, enters with content, but not with formal structure and dynamics. Once we specify what A'ing or refraining from A'ing are, what exactly the agent desires and intends, what the context of a given instance of agency is, and perhaps who the agent is, it becomes possible to evaluate any instance of self-trust, self-distrust or self-mistrust from a moral point of view. So, for example, we (and perhaps the agent himself) would judge it morally good for someone who tends to weaken at the sight of blood to distrust himself to stop the bleeding of a seriously injured party; his weakness could result in the injured party bleeding to death, so he shouldn't even form the intention to stop the bleeding much less try to do so, even though he might strongly desire to do so. The terrorist's self-trust might help make her an efficient killing instrument, but (we assume) this is not a morally good thing even if it's for a good cause.<sup>30</sup> I note that the rationality requirements in self-trust for belief justification and reasonable expectations certainly allow for, but they do not themselves represent, moments of moral reflection on the agent's part.

3) Self-trust, as presented here, parallels a special case of interpersonal trust: 2 trustworthy people trusting each other to A or to refrain from A'ing. In game theory, this is the standard 2-person outcome-dominant Nash-equilibrium in the stag-hunt; if agents trust each other, they decide to cooperate or to continue to cooperate in the iterated stag-hunt. Self-distrust tracks the parallel case: 2 untrustworthy agents distrusting each other to A or to not-A. This is the risk-dominant Nash-equilibrium in the stag-hunt and the maximin rational choice; if agents distrust each other, it is rational to defect. These game-theoretic models can help the analysis of intrapersonal trust and distrust because of the partial analogy between intra- and interpersonal trust. Such areas of analogy, however, are generated by the metaphysical theory of the self or person as an aggregate or composite of time-distributed sub-selves, the accumulation of stages of an event. The assistance, then, is purchased at the cost of opening the above analysis to all the philosophical controversy and criticism to which this theory of the self is open, especially those concerning personal unity and identity through time.

4) Self-trust and self-distrust, by condition (iii), are seen to depend on the agent's autonomy and rationality. If a later self couldn't believe that its earlier self was both autonomous and reasonable in placing it under a normative expectation, it surely couldn't trust that earlier self and that earlier self would not be trustworthy in its eyes. Govier (1993, p. 111), however, argues that the relation between self-trust and (procedural) autonomy is the reverse; she argues that an agent's autonomy depends on self-trust. I believe Govier is too quick here; an agent's autonomy is not the same as an agent acting with autonomy, just as deciding is not the same as acting on one's decision. Self-trust might well be necessary for acting with (and perhaps even without) autonomy, but autonomously (and reasonably) placing one's self under a normative expectation is not, in the usual sense, an action – a behavior; it is more in the nature of inwardly exercising the right of self-government. Thus, I don't see my analysis of self-trust clashing, in this respect, with that of Govier.

Importantly, the rationality constraint in condition (iii) is epistemic not practical; it concerns evidence for a belief, not deliberation for a decision. Thus, there is a stage of practical self-trust that imports epistemic norms for belief, but not for truth; that is, norms for justifying: (a) a future self's belief that its former self, whose "rule" it agrees to live by in the undertaking at issue, is trustworthy in its expectations for that future self, and (b) an earlier self's belief that its future self is sufficiently capable to carry out its intention to A or refrain from A'ing while confronted by countervailing forces. A self-trusting agent, just as much as an other-trusting agent, should not be epistemically gullible or biased; one would expect an important place for epistemic norms of self-examination and self-criticism within the dynamics of self-trust. The above analysis allows for this.

5) Hinchman (2003) sees self-trust as the "rational link" between intending and acting. The problem he addresses is: why should (not: why does) an agent carry out (be "intention-executing" in his terminology) an intention formed by its former self? That is: what reason does the agent have at the time of execution to execute its earlier intention, given that there is no reason not to. It can't be, for example, because at the time of action the agent re-deliberates and reaffirms its earlier intention; for Hinchman this just moves the same problem forward in time, as does claiming the explanation lies in intention-constancy by way of a memory at the time of action of the agent's earlier intention. Hinchman argues that self-trust is the solution; a relationship of trust runs from the later intention-executing self

back to the earlier intention-forming self, and this trust functions as a “pre-emptive reason” to act on the intention, providing the agent’s later self judges its earlier self to be trustworthy.<sup>31</sup> This self-trust is essential to rational practical agency for without it, in Hinchman’s view, there looms an un-bridged gap of reason between intention and action.

There are two points of disagreement between Hinchman’s and my analysis of practical self-trust. First, I see self-trust as rationally bridging the space between a practical agent’s desire to do or refrain from doing something and the agent’s intention to do or to refrain from doing it. Why should an agent who desires to do something form an intention to do it? Why not just live with the desire? One explanation might be self-distrust; a self-distrusting agent does just that, or when possible seeks external assistance strategies, for she believes (rightly or wrongly) – prior to forming an intention – that she does not have what it takes to do what she desires to do. If there is a form of practical self-trust that links intention with action – and Hinchman’s analysis convincingly shows that there is – it is further downstream in practical agency and would have to presuppose the earlier self-trust that I have analyzed here: that leading to forming an intention in the first place to do what the agent desires to do. Self-trust, as I see it, first (because practical agency typically begins with a desire) and foremost (because practical agency is primarily a matter of the power and limits of practical reason) is a link between a desire to do something and an intention or decision to do it – an intention formed with sufficient belief in one’s self and resolve that the agent has little room to fret, hesitate, or re-consider either when forming it or when the time comes for carrying it out. And, this is why self-trust is a sometimes dangerous thing.

The second disagreement between Hinchman’s conception of self-trust and the one suggested here is that for Hinchman the trust seems to run in one direction: from a later self at the time of action to an earlier intending self whom that later self believes (with good reason) to be trustworthy (see note 31). But isn’t this just part the picture? It would seem that a person’s earlier self, in believing that it should appear – and be – trustworthy to its future self, must to a degree still trust that future self to then act (as intended) on the reason of trustworthiness and not sink into “pathological” hesitation, forgetfulness, procrastination, self-doubt, self-non-confidence or over-scrupulosity. This is the case, I believe, because even self-trust as a “pre-emptive reason” is no guarantee. Nor should it be; it would be deeply troubling if an agent were so bound by its self-trust that it *couldn’t* freely, even whimsically, at the very last moment of action spontaneously ignore her earlier intention, rebel against it, or give up for no good reason. This may be irrational (even though externally fortunate should a last minute whimsical opt-out means a great harm doesn’t get done), and perhaps frightening to an agent’s earlier self who then must be deeply uncertain and always vulnerable to risk even about self-trust – about the kind of person s/he will discover themselves to be in some future circumstance – but we wouldn’t accept an account of self-trust that made it impossible. The trust that runs from the earlier intending to the later executing self would acknowledge trust-risk: that the agent’s freedom to “change my mind” has not disappeared, even at the very last moment of action, nor the possibility to “surprise myself” in acting. It would seem that some forward self-trust, then, is always needed; an agent’s future intention-executing self ought to appear trustworthy to the earlier self, as much as its earlier intending self appears trustworthy to the self that is about to act. Hinchman’s account of self-trust, in other words, gives us deep retrospective insights about agency: how in a unit of practical agency the reasonable agent is, so to speak, “pushed”

to act by its self-trust. Yet, at least intuitively, practical agency is also – if not primarily – a future oriented rationality: an agent also needs to be “pulled” in its journey by its self-trust. Granted, the problem on which Hinchman sets his sights is inherently backward looking, so the unevenness of his account (granting that it is) may well be justified by the bounds of his topic. Still, the account of self-trust offered here provides, I believe, a fuller picture than Hinchman’s.

##### 5. Some connections and some distinctions between self-trust and self-confidence

Based on the analyses of self-confidence and self-trust offered here:

(1) what are some of the ways in which these two intrapersonal relations are importantly different for the exercise of practical agency? Also,

(2) how are these two intrapersonal relations conceptually related?

(1) First, self-confidence depends primarily on an agent’s belief as to how well the agent’s action will be performed and be judged, given the expectations of others. The self-confident agent intends to (or must) do something, desires to do it well, believes she will, and expects to receive recognition for satisfying norms of performance not only her own. An agent’s self-confidence or lack thereof, then, is about the agent’s *performance*; the agent’s beliefs about the agent’s body – e.g., voice and speech, bodily actions and appearance, how the agent looks, and what results from these – and beliefs about how it will be perceived and judged by others should the agent perform the intended action, are integral to self-confidence or its lack. Self-confidence is only indirectly about the person: to the degree, perhaps, that a performance would be evaluated as reflecting qualities of the person who performs such as dedication to prior preparation, training, practice, seriousness of effort, etc. For most actions done with self-confidence, however, doing well or poorly says little or nothing about the character or the inner qualities, much less the value, of the person. The exception might be the case of the overly self-confident or the overly self-non-confident agent, mistakes in self-knowledge that are typically judged to be flaws in the person. For the most part, the self-confident (or opposite) agent believes s/he is being judged *qua* performer, not *qua* person; in failing to act well, the self-confident agent believes she primarily lets others down regarding their expectations about her actions, who then think less of her as a performer but not (or at least shouldn’t) as a person. This dimension of the “eyes of others” and “public expectations” the self-confident or self-non-confident agent believes she faces is the basis, I suggest, for any emotional or physiological states that typically accompany these self-relations and their potential beneficial or harmful effect on the agent’s performance. It is reasonable, for example, for the coach and for the athlete both to be concerned about the athlete’s self-confidence or lack of self-confidence in a future competitive sports event, and in my (granted somewhat limited) experience they are, only if they believe that these will affect the athlete’s performance.

In contrast, self-trust depends primarily on an agent’s self-evaluation regarding the agent’s capability to satisfy a self-imposed normative expectation. The agent desires to do something, believes this takes being a certain kind of person (*qua* practical agent), expects of herself that she ought to be that kind of person, believes that she is, and consequently forms an intention to do what she wants to do. Self-trust,

then, is primarily about how one views one's self and the kind of person one takes one's self to be: namely, capable and trustworthy in regard not just to a given undertaking, but also in the domain in which a "moral strength" – a virtue – *might* be exercised. Self-trust, then, is not just about the agent's specifically intended action; it is about enduring practical capabilities, within and beyond any particular unit of agency. In failing to act as intended, the self-trusting agent primarily lets himself down; self-disappointment is the important result and only incidentally the disappointment of others. This aspect of "in my own eyes" central to self-trust and self-distrust, in contrast to the "eyes of others" central to self-confidence and self-non-confidence, I suggest, is a step toward explaining the lack of similar emotional and physiological states in the case of self-trust/distrust that we typically associate with self-confidence/non-confidence. In my experience, neither the coach nor the athlete, for example, show concern about the athlete's self-trust or self-distrust in a sports event; if this analysis is correct, they shouldn't if it's the athlete's performance and not her character that matters.

A second distinction is that self-confidence does not itself contain normativity; it involves an agent's intentions, desires, and beliefs. There is no action that an agent can intend to do that the agent can't be self-confident about doing (or lack self-confidence about doing). The norms of performance the agent is challenged to satisfy and is confident about meeting are necessary as "public" and "external" to the agent; but even in the case where the agent alone applies them to evaluate his performance they are "internally" part of – constitutive of – the agent's self-confidence as an "external" expectation. An agent's self-confidence, then, might be summed up by the words "I will show *them* I can do it" – the self-confident agent is not trying to prove something to herself.

Self-trust, however, contains normativity: the ideals of autonomy, reasonable expectations, belief justification, and trustworthiness are integral to self-trust. These serve to constrain an agent's self-trust; there are many actions an agent might desire to do that the agent can't be self-trusting about doing, and many the agent should be self-distrusting about doing. Self-trust, perhaps, can be summed up by the words "I should be and can be believed in, by *myself*, to do it."

A third distinction is that, typically, it is commendable but not necessary to the exercise of effective practical agency that the agent desires to do something well, or even that it be done well, just that it be done sufficiently to achieve its end even though this might be judged a mediocre performance.<sup>32</sup> In fact, there are no norms of performance for many (most?) actions; and, there are human activities that apparently don't even have norms of completion, such as counting the natural Arabic numerals, parenting one's child, or appreciating a great work of art. It follows that in such cases self-confidence plays no role in successful practical agency. Self-trust, however, is an essential part of successful practical agency of an important form. Self-trust and self-distrust operate between the  $t_1$  point of desire and the  $t_2$  point of forming an intention in the general structure of a unit of practical agency described in part I above. A desire to do something (whether morally good or bad) can't reach the point of forming an intention to do it without self-trust and self-distrust will cut it off, or so I've argued. And without an intention to do something, it is hard to imagine an agent acting successfully on desire alone; at any rate, it certainly would not qualify as rational practical agency, and without a lot of luck wouldn't be successful action. Without self-trust, then, an important form of practical agency – though not all practical agency – breaks down in its earliest stage.

There are several more distinctions between self-confidence and self-trust that follows from the above analyses:

- that self-confidence tends to eliminate the agent's risk aversion and sense of vulnerability even in high-stakes performances, but self-trust remains open to a sense of risk and vulnerability,
- that even though both self-confidence and self-trust are intrapersonal relations, neither self-confidence nor its lack can exist in a universe containing only one person whereas self-trust and self-distrust can,
- that while both self-confidence and self-trust are beliefs, and as such share the basic propositional-attitude structure: subject-content-object, the general content and object of each type of belief are importantly different, and they are not subject to the same epistemic norms.

However, the first three argued points of contrast are sufficient, I believe, to show that these two intrapersonal relations, though they might have some structural and dynamic similarities, turn out to be quite different aspects of practical agency. It follows that "self-confidence" and "self-trust" are not just two ways of referring to the same thing about a person, even if they might be used this way in everyday language.

2) How are self-confidence and self-trust, as these are defined above, conceptually – as opposed to empirically or causally – related? They are logically compatible if it is possible to imagine one practical agent in one unit of agency of whom both sets of conditions are true – that is, a practical agent who is both self-trusting and self-confident with respect to the same action – and incompatible if this is not possible. It is easy to show that they are compatible.

Imagine practical agent Jill (=S) who desires to exercise well (=A), say for reasons of health, and who believes that to exercise well she must have a strong, long-term commitment to good health (=X). So: let S at  $t_1$  desire to A at  $t_3$ , and let S at  $t_2$  intend to A at  $t_3$ , and let S perform A at  $t_3$ . In the interval between  $t_1$  and  $t_2$  let S believe that capacity X is required and productive of A'ing, and let S normatively expect that S has X at  $t_3$ , as a result of which S forms her  $t_2$  intention to A at  $t_3$ . And in the interval between  $t_2$  and  $t_3$  let S believe that to A is to be evaluated by public norms of performance (e.g. an exercise coach or readouts on an exercise machine) as having A'ed, and let S desire to meet these norms of performance at  $t_3$ , and finally let S believe that at  $t_3$  she does A according to the norms of A'ing. Whether we further imagine S A'ing or not A'ing at  $t_3$ , this is so far a possible scenario in which all the conditions for self-trust and for self-confidence are true. Thus, self-trust and self-confidence are compatible; a practical agent can be both within one unit of agency. I note, however, that this model does not show that this imaginary agent is either trusting about her self-confidence or confident about her self-trust, only that it is possible to be both self-trusting and self-confident within a single practical undertaking.

Self-distrust and self-non-confidence, however, are mutually exclusive; it is impossible for an agent to be both with respect to the same practical undertaking, and the presence of one implies the absence of the other. We see this in the contradiction between condition (i) for lacking self-confidence (an agent must form an intention to A: the action about which the agent lacks self-confidence) and condition ( $v_a$ ) for

self-distrust (after satisfying conditions (i)-(iv<sub>a</sub>) the agent does not form an intention to A). For any given action A, an agent cannot both form and not form an intention to A in a single unit of agency; thus, an agent cannot be both self-distrusting and self-non-confident – a self-distrusting agent does not reach the point in practical agency for a lack of self-confidence to take effect, thus an agent lacking in self-confidence can't also be self-distrusting.<sup>33</sup>

Self-trust and self-confidence are logically independent if it is possible for a practical agent to satisfy one set of conditions without satisfying the other. So, is it possible to imagine: (a) a self-trusting agent who has no self-confidence, and (b) a self-confidence agent who has no self-trust?

(a) While it may be part of human psychology for a self-trusting person to tend to have self-confidence, the issue here is whether conceptually self-trust implies self-confidence. Clearly it does not. Imagine Jack who desires to swim back to shore against an increasingly strong ocean undertow after swimming a bit too far from shore. Jack has no desire to swim well and doesn't care about being judged a poor swimmer, and does not believe that his efforts will be evaluated; he simply wants to make it back to shore. He believes this will take determination and strength of effort, and the ability to fight panic and stay focused in case the undertow becomes stronger and he should start to tire. He normatively expects that, as a swimmer in his situation, he ought to have such determination and ability, and personally believes that he does; he is trustworthy in his own eyes to make it back and not to give up. As a result he forms the intention to swim back to shore and begins the effort. All of the conditions for self-trust are true of this agent's undertaking, but only the first condition for self-confidence, forming an intention to perform an action, is true; the others are false of Jack. Therefore, this practical agent trusts himself to swim back to shore, but is not confident about his performance (which is not to say that Jack is confident that he won't make it). It follows that for a person within a given unity of practical agency, self-trust does not imply self-confidence.

(b) If an agent is self-confident, must the agent be self-trusting? Imagine Jill is asked by an elderly neighbor for a small favor: drive to the local pharmacy and pick up her medication. Jill is neutral about the favor; she has no desire to drive to the pharmacy and no desire not to, no desire to help her neighbor and no desire not to. Jill agrees and says she'll do (Jill forms the intention to do) the errand in an hour. The neighbor is grateful. Jill believes that there are norms for good driving involving speed, safety, road courtesy, etc. and norms of doing favors and neighborliness. Given her intention, Jill desires to drive to the pharmacy and back in a way that satisfies these norms of good driving and neighborliness, and would like to be judged as having done so. Finally, Jill believes that in an hour she will drive and complete the favor according to these norms; she believes that she will be a good driver and a good neighbor in this easy task she has agreed to do. In other words, Jill has complete self-confidence in carrying out her intention well and being a good neighbor in doing this small favor, yet self-trust does not enter the picture. All the conditions on self-trust fail to hold in this scenario, and all the conditions on self-confidence are satisfied. Thus, self-confidence does not imply self-trust.

We see that self-trust and self-confidence do not imply each other and are, therefore, conceptually independent. What of self-distrust and a lack of self-confidence (self-non-confidence)? Given that these two self-relations are incompatible, as argued above, it follows that an agent who is self-distrusting can't also be self-non-confident. Scenario (c) offers an example of a self-distrusting agent who has no lack self-confidence. I note that not lacking self-confidence is not equivalent to having self-confidence; an agent could do something easy and automatically, for example get out of bed in the morning, without any self-confidence and without any self-non-confidence (without a lack of self-confidence).

(c) Imagine Jack who desires to continue driving late at night in order to reach home. He feels tired and sleepy. He believes that it will take special effort to keep from falling asleep at the wheel, and he normatively expects that a driver in his situation ought to be capable of keeping awake and fight-off such drowsiness. But Jack believes he does not have the capability to do this; he distrusts himself to keep awake while wishing that he was the kind of person who would not be untrustworthy in such situations. Jack has no "external assistance" to turn to, to make up for his perceived agential weakness: no drugs, for example, or no external source of sensory stimulation. As a result, Jack forms no intention to keep driving home; instead, he forms the intention to pull off the road into the nearest rest-stop. All the conditions on self-distrust are true of Jack in this scenario, but none of the conditions on lacking self-confidence hold.

Given that self-distrust and self-non-confidence are mutually exclusive, it follows that an agent who lacks self-confidence can't also be self-distrusting within the same unit of agency. Scenario (d) provides an example. I note that not being self-distrusting is not equivalent to being self-trusting; an agent could do something easy and automatically, for example get into bed at night, without either self-trust or self-distrust.

(d) Jill, a recent college graduate who still lives at home, believes her mother (a dominant figure in Jill's life who worries about Jill's life-style and future) is overly-critical of her choices, actions, and "lack of long-term planning." One day Jill's mother asks her to do some food shopping at the local market. Jill has no desire to do this simple task, and believes that doing it takes no special capability or strength that Jill normatively expects of herself and further believes that she possesses or fails to possess. Thus, neither self-distrust nor self-trust operates in Jill's unit of agency: shop for food. Feeling obligated to accept her mother's request, Jill agrees and forms the intention to go food shopping. On the way to the market, Jill both desires to fulfill her mother's request in a way that leaves no cause for criticism from her mother, and believes that no matter what she purchases, her mother will still find fault and be critical of Jill's "performance." Jill experiences a lack of self-confidence about carrying out her intention to shop for food in the way she desires; that is, in a way that satisfies her mother's (perhaps perfectionist) norms and expectations. In this scenario all the conditions for lacking self-confidence are met by Jill, but none of the conditions for self-distrust hold.



No doubt, in the realm of human psychology self-trust and self-confidence are connected in complex correlational and causal ways. The same is probably true of self-distrust and the lack of self-confidence. But these are topics for empirical research. Given the analyses of these four intrapersonal relations above, it follows that self-trust and self-confidence are conceptually independent and compatible; however, self-distrust and self-non-confidence are not independent: the presence of one is sufficient for the absence of the other. As to the conceptual relation between self-trust and self-distrust, and between self-confidence and self-non-confidence, it follows from the above analysis that it is not possible for an agent, in the same unit of agency, to be both but it is possible for an agent in a given unit of agency to be neither.<sup>34</sup>

---

#### Notes:

1. See Davidson (2001) "How is weakness of will possible" and (2004) essays 11-14 on problems of agential irrationality.
2. Govier (1993)
3. Elster (2000) Footnote 9, pp. 5-6.
4. See: Ainslie (2001) especially chapter 3 and part II ;Elster (2000) especially I.1, I.3, I.4, II.8; Schelling (1984) especially p. 57ff.
5. Hinchman (2003).
6. This is a purely pragmatic decision; I find exploring intrapersonal relations conceptually easier within the second model of the self than within the first with respect to the layering of parts, their process of building up and development. I don't mean to claim that I find the second model true and the first false, or even that it is more likely true than the first, although – given that we live our lives in time – adding the dynamic dimension to intrapersonal relations does seem to add realism.
7. Lehrer (1997, ch. 1) argues that this kind of self-reference – trusting oneself as worthy of one's self-trust , and trusting oneself as worthy of this, ..., – is a valuable self-justification of the explanation why a person requires self-trust; he does not see self-reference in the capacity for self-trust as potentially insulating a person from, or depriving a person of, the benefits of and the need for self-doubt and self-criticism in epistemic and practical reasoning – no doubt because such self-doubt and self-criticism would have to be trusted to be effective.
8. See Lehrer (1997), Hinchman (2003), and Govier (1993). Govier does point out the negative side of self-trust, but only for the case in which it is unfounded and excessive (p. 115), and Hinchman considers self-trust a good thing and bad ("pathological") only when it is unreasonably withheld (pp. 32-33).
9. Govier (1993) p. 110ff; her argument, in schematic form, that self-trust is a necessary condition for self-respect (p. 110) runs as follows: P1 - If a person experiences certain kinds of self-doubt, then that person has a lack of self-respect. P2 – Thus, if a person has self-respect, then the person is not in (those kinds of) self-doubt. P3 – A person is not in (those kinds of) self-doubt iff the person has self-trust. Therefore, if a person has self-respect, then the person has self-trust. The argument is valid. A similar argument concludes that self-trust is a necessary condition for what Govier, following Meyers, calls "procedural autonomy."

10. Lehrer (1997) takes issue with the image of “foundation” and prefers the striking image of self-trust as the circular “keystone” in the dome of an epistemic and practical “building” whose arches both form and support its dome. (p. 10ff).
11. Hinchman (2003).
12. Govier (1993) p. 115. Consider another somewhat representative passage in which self-trust and self-confidence appear to be either equated or confused. “One must confidently depend on oneself to think accurately, deliberate reasonably, make sound decisions, carry out sensible plans and implement worthy goals. Without core self-trust, a person is so open to the manipulations of others as to lose any sense of a meaningful agency.” (p. 115). Elsewhere (pp. 108 and 112) self-distrust is equated with lack of self-confidence.
13. Hinchman (2003), p. 27. Another place where self-confidence and self-trust are brought dangerously close, if not equated, is in footnote 14. Hinchman writes, “Velleman considers a case in which an agent loses confidence in herself ... that she will not stick to her plans: ... that she will not follow through on her plans begins to shake her faith in herself as a planner. ... This is, I concede, a failure of self-trust... .” (p. 47). Hinchman goes on to explain that this kind of failure of self-trust, however, is not the kind that he is addressing in the body of his article.
14. This would be the case even if an agent intended to do this: an action that didn’t match the action the agent intended to do. You intend to do A, someone now counts on you to do A, you wish to deceive this person, so you intend to do something other (it doesn’t matter what) than A, your intended action.
15. Presumably S has a reason – perhaps a prior decision or commitment, an obligation, a promise, a contract, or a goal – at  $t_1$  for making an “appointment” for that part of her future that will be taken up fulfilling her former self’s intention to A at  $t_2$ . But exploring S’s possible good or bad reasons for intending to A is beyond the range of my topic.
16. Evaluated when? Independent of S’s belief, the variations are interesting to consider: perhaps simultaneous with S’s A’ing, perhaps long after S does A, perhaps never, and perhaps even before S performs A as a type of predictive or provisional pre-evaluation (thought this could be unfair to S). Here is a case in which self-confidence might be thought to run from future (or present) to past, instead of the typical past (or present) to future direction. Suppose at  $t_1$  S intends to A at  $t_2$ , and at  $t_2$  S does A. But S’s  $t_2$  performance is not evaluated until  $t_4$ . At  $t_3$  S is waiting for a future evaluation of an action completed in S’s past. If S is self-confident about A’ing well, then at  $t_3$  S is self-confident about S’s  $t_2$  action (and presumably has been self-confident about A’ing from  $t_1$  to  $t_3$ ). Note, however, that this variation is derivative; it seems acceptable only because something remains open and unfinished about S’s A’ing at  $t_2$ : the “jury is still out.” For past actions that are over-and-done-with, it is hard to see in what sense the agent could have self-confidence or lack self-confidence about them; it would be retrospective pride perhaps, or regret. For practical agency, the future is much more important than the past.
17. Schelling (1984). “My suggestion is that we get help from comparing self-management with the way one tries to manage another, another who is in a special relation to one’s self. Many of the skills and maxims and stratagems for coping with one’s own behavior become less mystifying and more familiar if we can recognize them as the same principles and stratagems that apply in managing someone else... .” p. 63.
18. Govier (1993), p. 104. In her analysis of self-trust, Govier does not appeal to the concept of a person as a collection of earlier and later selves, her framework is more that of a person as a collection of diverse interests or diverse units of epistemic and practical capacities and achievements. Nevertheless, she uses the strategy of modeling self-trust as analogous to interpersonal trust.
19. Hinchman (2003), p. 26.
20. See Carr (2012). To continue the terminology set up in that article, I will here use “distrust” for the opposite of trust, and “self-distrust” as the opposite of self-trust. I reserve “mistrust,” and thus “self-mistrust,” for instances of trust and self-trust that are out of line with the deserved trustworthiness of the trusted (the trustee). So, “mistrust” is trusting either too little or too much (it is trust that misses the mark, as it were) whereas “distrust” will here mean more than just an absence of trust or a withholding of trust (perhaps for reasons having nothing to do with the trusted’s trustworthiness), and more than mistrust; it

means an intentional decision or attitude not to trust, typically based on the belief that someone has about another person (or in the case of self-distrust, a person has about himself) that s/he can't be trusted and should not be trusted because s/he is untrustworthy. Also, I will assume that trust and self-trust, trustworthiness and self-trustworthiness, and the related concepts of mistrust and distrust, have intensities; that they exist in degrees or amounts so that it makes sense to use "more" and "less" and "equal degrees" as descriptors. "Too much trust" as a case of mistrust, then, would not be the same as "being too trusting" – that is, being gullible and ready to trust without reserve.

21. Hinchman (2003) is sensitive to an important disanalogy between trusting others and self-trust: in the interpersonal case, he argues, trust does not imply compliance, but self-trust reduces to compliance due to the authority the earlier trusted self has over the later trusting self. For a later self to question or doubt or resist or rebel against this authority without good reason would be to destroy practical agency (p. 34).

22. Schelling (1984) provides a variety of immediately recognizable examples. While he doesn't use the trust-distrust terminology to examine intrapersonal management, it is easy to see that many if not most of his examples fall under these categories.

23. The assumption of course is that desires are motivating, but this does not exclude reasons, decisions or intentions from also being understood as motivating actions.

24. This phenomenological argument is quite weak though, I believe, not wrong. There seems also to be a place for a kind self-trust in an agent who intends to A and who lacks any prior desire one way or the other about A'ing (we can imaginatively put ourselves in such an agent's shoes), but not for the agent who desires to A but does not intend to A, or perhaps even intends to not-A – this would be the place for self-distrust, if anything. However, the kind of self-trust that operates between intention and action, assuming it is not a form of self-confidence, would not be about the (strength of) ability to A or refrain from A'ing, e.g. the ability to resist stealing money when the need and opportunity are present), for intending implies the belief in ability. Hinchman's (2003) analysis of self-trust locates it between intention and action with no requirement that the agent have a prior desire to do what the agent intends to do. But Hinchman's aim is to see in self-trust the source of reason to A, given the prior intention to A; my aim is to distinguish self-trust from self-confidence, and for my aim the self-trust that leads to forming an intention to A (and the self-distrust that leads to giving up even forming an intention to A) more clearly serves.

25. The self-evaluation would still be the case in the event that an agent accepts another person's judgment that the agent should possess the ability, skill, resolve or strength of effort at issue, as long as "accepting" means "makes it her own" or "applies it to one's self".

26. Perhaps an example would be: a drug addict desires a fix, and believes he should not have the ability to get it (because he also desires to break his addiction). But here, the normative belief lines up with second desire, not with the first, with which it conflicts. Even drug addicts are coherent practical agents, or they would have a lot more to worry about than their addiction.

27. The argument here is standard though not without controversy: (1) to desire the end is to desire the means, and (2) the object of desire is valuable to the desirer, therefore (3) to desire to A or to refrain from A'ing is to value the capability to A or refrain from A'ing.

28. The agent's reasoning or set of beliefs will perhaps run as follows: (1) Ideally, a practical agent of that kind has this ability. (2) I am such a practical agent, or aspire to be. (3) I should have this ability.

29. Both Schelling (1984) and Elster (2000) prefer "imperfectly rational" to "irrational" as the better description for failures to satisfy the ideals of practical rationality.

30. Compare Govier's influential analysis (1993) in which (accurate) self-trust is argued not only essential to effective practical agency, it is presented as central for having intrinsic moral self-worth (p. 114ff). Self-distrust is, by parallel argument, not only destructive of practical agency, it is destructive of essential parts of an agent's humanity. Hinchman (2003) likewise argues for the centrality of self-trust to practical agency and self-distrust to its breakdown, but his analysis is more formal and does not address directly the question of its intrinsic goodness as does Govier. Lehrer (1997) links self-trust and trustworthiness to the

worth of practical and epistemic goals in a way that seems to make immoral self-trust impossible: "I am not worthy of my trust in my pursuits unless the ends are worthy of my trust, and the means by which I pursue them are worthy of my trust as well." (p. 56ff).

31. Hinchman (2003). "Rather, the earlier self must believe that the later self will follow through on the intention *through an exercise of reasonable trust*. And, as I've now argued, the earlier self must believe that it will appear trustworthy on the matter at hand." (p. 35ff).

32. I qualify this claim with "typically" because an agent's intention might be to gain self-confidence itself, or more of it. Presumably, this intention would be based on an agent's belief that he lacked self-confidence or didn't have enough of it for an area of activity of interest to the agent. But for the most part practical agency is not exercised for self-improvement.

33. It might be objected that a self-distrusting agent can nevertheless still form an intention to do an action that requires a capability the agent believes she does not possess; isn't this at least possible? No, given that the self-distrust in question is rational; by condition (iii) for self-distrust, the agent satisfies certain rationality constraints that makes the self-distrusting agent's untrustworthiness justified. Such an ideal agent could not ignore, forget or override her self-distrust and proceed to form an intention to A, as if it didn't count.

34. I am grateful to my colleagues, Jerry Dolan and Herman Tavani, for several penetrating comments and suggestions.

---

## References

Ainslie, George. 2001. *Breakdown of the Will*. Cambridge University Press

Carr, Lloyd. 2012. "Trust: an analysis of some aspects." Available at: <http://www.rivier.edu/faculty/lcarr>

Davidson, D. 2001. "How is weakness of will possible?" *Essays on Actions and Events*. Oxford University Press.

\_\_\_\_\_, 2004. Part 3 Essays on "Irrationality." *Problems of Rationality*. Oxford University Press

Elster, Jon. 2000. *Ulysses Unbound*. Cambridge University Press.

Govier, Trudy. 1993. "Self-Trust, Autonomy, and Self-Esteem." *Hypatia* vol.8, no.1, 99-120.

Hinchman, Edward. 2003. "Trust and Diachronic Agency." *Nous* 37:1, 25-51

Lehrer, Keith. 1997. *Self-Trust: A Study of Reason, Knowledge and Autonomy*. Oxford University Press

Schelling, Thomas. 1984. "The Intimate Contest for Self-Command." *Choice and Consequence*. Harvard University Press

\_\_\_\_\_. 1984. "Ethics, Law, and the Exercise of Self-Command." *Choice and Consequence*. Harvard University Press