Hume, Malebranche, and the Self-Justification of the Passions
Éléonore Le Jallé
Hume Studies Volume 38, Number 2 (2012), 201-220.

Hume Studies
Volume 38, Number 2, 2012, pp. 201–220

Hume, Malebranche, and the Self-Justification of the Passions

ÉLÉONORE LE JALLÉ

Abstract: In Book 2 of the *Treatise*, Hume echoes Malebranche's *Search after Truth* in noticing that all our passions tend to justify themselves. I reveal this borrowing and examine how this phenomenon of the self-justification of the passions fits into Malebranche's and Hume's different approaches to the passions and their links to judgment and truth. I maintain that whereas Malebranche aims to warn against the errors that passions involve when justifying themselves, Hume only considers the self-justification of the passions as an example of the dynamic nature of the mind, other examples of which are displayed elsewhere in the *Treatise*. I also show that Hume's understanding of the self-justification principle illustrates another important theme of Books 1 and 2, namely, the reciprocal influences of the imagination and the passions.

The Physiological Library's catalogue shows that Hume had access to Malebranche's sixth edition of *De la recherche de la vérité* while a student in Edinburgh.[1] The *Recherche* is also included in the David Hume's Library.[2] While Hume did not agree with Malebranche on all things, a number of commentators have argued that Hume borrowed many points from Malebranche, not only concerning causality and the famous example of the billiard balls[3] but also on other subjects. Charles McCracken's *Malebranche and British Philosophy* compares Hume and Malebranche on causation, self-knowledge, kinds of truth, the relation between imagination and belief, and ethics.[4] Peter Kail has more recently devoted three articles to the connection

Eléonore Le Jallé, Department of Philosophy, UMR 8163 (STL), University Lille 3, Domaine Universitaire du Pont-de-Bois, BP 60149, 59653 Villeneuve d'Ascq, France.
Email: e.lejalle@free.fr.

between Hume and Malebranche, one on ethics, one on causation and the self, and one on rationalism.[5] Concerning the passions, which will be the topic of the present article, in a recent article, Susan James discusses Malebranche's and Hume's explanations of respect and contempt, focusing on their similarities but stressing, at the same time, their different reasons for considering these topics.[6] As she writes, "Hume's account of the passions echoes and engages with [Malebranche's] *De la Recherche*" (James, "Sympathy and Comparison," 107). The Norton and Norton edition of the *Treatise* indicates many allusions to Malebranche or borrowings from him, some of which relate to the passions.[7] However, a particular borrowing seems to have escaped the commentators: Hume echoes Malebranche in noticing that all our passions tend to justify themselves by causing some judgments which foster them. To give just one example of this phenomenon, which can be found in both Malebranche and Hume, if we hate a person (for instance, our antagonist in a law-suit), we are apt to think that he has bad intentions towards us because this opinion justifies our hatred and thus preserves it.

In this paper I establish this borrowing and examine how this phenomenon of *the self-justification of the passions* fits into Malebranche's and Hume's different views about the passions and their links to judgments and truth. Thus, my intention is twofold. On the one hand, I shall argue that Hume has directly borrowed the self-justification principle from Malebranche. This principle can be defined as follows: all the passions seek their own justification to avoid diminution (or preserve their agitation) and thus produce opinions which foster them. On the other hand, I intend to show that whereas Malebranche aims to warn against the errors that passions engender when justifying themselves, Hume only considers the self-justification of the passions as an expression of a certain dynamics of the mind which is discussed elsewhere in the *Treatise*. In addition, I shall link Hume's self-justification principle with other themes of the *Treatise*, especially that of the reciprocal influences of the imagination and the passions. However, before focusing on this particular principle and its significance for Malebranche and Hume, a more general picture of their respective approaches to the passions is needed.

As I shall spell out three important differences between Malebranche's and Hume's approaches to the passions in the first section of this article, I will now only sketch the general intents of their respective projects. On a superficial level, Malebranche's and Hume's considerations on the passions might seem to be part of a common psychological project, namely to attain knowledge of the human mind, which Malebranche depicts as "the most beautiful, the most pleasant, and the most necessary of all knowledge," inasmuch as "of all the human sciences, the science of man is the most worthy."[8] As David and Mary Norton remark in an annotation to Hume's introduction to *A Treatise of Human Nature*, Malebranche belongs to "many of Hume's predecessors [who] recommended the careful study

of human nature" as the most worthy of all the sciences. Nevertheless, as the Nortons also add, Malebranche pursues "a significantly different agenda."[9] Indeed, Malebranche's intention of entirely explaining the nature of the human mind explicitly depends on a more specific intention: revealing the errors of the human mind, which are caused by its union with a body. Explaining how the mind works thus enables Malebranche to discover the general causes of its errors. The passions, as well as the senses and the imagination, are explicitly involved in this latter project, insofar as they cause a false and confused knowledge: "they dazzle us and seduce us in every instance" (*Search after Truth*, preface, xxxix). By contrast, Hume's explanation of the passions depends on an overall psychological project which is not, at the same time, a normative one. Explaining the passions is part and parcel of Hume's ambition to explain all the perceptions of the mind. Hume himself defines the passions as "secondary, or reflective impressions" proceeding from impressions of sensations "either immediately or by the interposition of [their] idea[s]" (T 2.1.1.1; SBN 275).[10] It is an important result of Hume's own science of human nature that any qualification of our passions *as such* as "unreasonable" (that is, "contrary to truth or reason") or "reasonable" is improper and unphilosophical (T 2.3.3.6; SBN 416). Thus, Hume's plainly descriptive position regarding the passions seems deeply rooted in his own opposition to the "method of thinking [of] the greatest part of moral philosophy, antient and modern," which consists in exalting a "suppos'd pre-eminence of reason," with its "eternity, invariableness, and divine origin," above passion, characterized by its "blindness, unconstancy, and deceitfulness" (T 2.3.3.1; SBN 413). There is little doubt that Malebranche's explicit claims against the deceitfulness of our passions made him, in Hume's eyes, an eminent supporter of this false moral philosophy. In addition, even though this is a more controversial point, I think that Hume's attack against reason is also a general theme in his *Treatise*, at least in parts 3 and 4 of Book 1 where skepticism is the issue. If this is the case, it would be Hume's more general attack on the pre-eminence of reason, and his correlative insistence on the imagination, that would motivate his rejection of Malebranche's rationalistic views. Nevertheless, as Malebranche's project also includes an overall description of the faculties and principles of the mind, it might also contain, in Hume's opinion, exact observations about its functioning. This explains why Hume borrowed many points from Malebranche's analysis of the passions, including the phenomenon of the self-justification of the passions, which is the topic of the present paper.

1. Hume and Malebranche on the Passions

Even though Hume borrows elements from Malebranche's analysis of the passions, three general types of differences can be found between their approaches:

differences in content, in method, and in their intentions, this third difference having already been hinted at in the preceding introduction.

A first difference concerns their classifications of the passions. Malebranche enumerates three basic passions: desire, joy, and sorrow. They all have good or evil as their objects, and they vary according to the fact that this good is either (a) possessed (*joy*), hoped to be possessed (*desire*), or (c) neither possessed nor hoped to be possessed, possessed without loss, or maintained (*sorrow*) (*Search after Truth*, 5.9; 391–92). According to Malebranche, all the passions relate to these three basic passions, and within a certain species, they vary according to the judgments made about good or evil (*Search after Truth*, 5.10; 396). Hume's division of the passions between "direct passions" (desire, aversion, grief, joy, hope, fear) and "indirect passions" (pride, humility, love, and hatred and their compositions) is obviously very different from Malebranche's classification. As James Fieser has argued in "Hume's Classification of the Passions and Its Precursors," the only link between Hume and the traditional Stoic classification, to which Malebranche is partly indebted, concerns the direct passions.[11] As far as the indirect passions are concerned, Hume totally innovates, and the main division between direct and indirect passions is also new. Conversely, the other Humean division of the passions between "calm" and "violent" is to be found in Malebranche and also in Hutcheson, who follows Malebranche on this point.[12] Nevertheless, an important difference must be noticed between Malebranche's and Hume's definitions of the violent passions. Whereas Malebranche associated the passions' different degrees of violence to an "infinite variation in the degree to which the spirits are agitated . . . and in their relation to the brain's fibers" (*Search after Truth*, 5.7; 374), Hume focuses on the degrees of "*emotion in the mind*," or on the presence or absence of a "disorder" and "sensible agitation" in the soul (T 2.3.3.8, 2.3.4.1; SBN 417, 419). Contrary to Malebranche, Hume does not speculate about the physiological causes of this emotivity, because the "emotion in the spirits," which he mentions in relation to violent passions, is for him identical to a "disorder" or "agitation in the mind" (T 2.3.4.5, 2.3.4.7; SBN 421).

A second difference between Malebranche and Hume concerns the use, or absence of use, of physiology. Malebranche often makes references to physiology in his explanations of the passions, beginning with their definition: "Here I call *passions* all the emotions that naturally affect the soul upon occasion of extraordinary motion in the animal spirits" (*Search after Truth*, 5.1; 337). According to Malebranche, this motion of the animal spirits is only the "natural or occasional" cause of the passions, while the will of God is their efficacious cause. It is "the continuous and all-powerful will of the Author of nature" which unites, for example, "the idea of an enemy's faults, or a passion of contempt or hatred, on the one hand, and the corporeal movement of the blood's parts striking against certain parts of the brain on the other" (*Search after Truth*, 5.1; 339). As for Hume,

he explains this very example of the hatred of an enemy's faults without any reference to the animal spirits or to the will of God, since he thinks that vice excites our hatred by the operation of a double relation of impressions and ideas. It is true that, in a number of places, as for example when Hume explains the swallowing of an inferior passion or emotion by a predominant passion (T 2.3.4.2; SBN 420),[13] Hume mentions an "agitation of the spirits" (T 2.3.4.2; SBN 420), which is, in my opinion, a sign or trace of Malebranche's text.[14] Nevertheless, Hume eventually identifies the "emotion in the spirits" or the "excitation of the spirits"[15] occurring in this explanation with an "*agitation in the mind*" (T 2.3.4.7; SBN 421, my emphasis).[16] Thus, Hume departs from Malebranche and provides an explanation of the *experience* we have of the "force" and "violence" of the passions, that is, of the "sensible agitation" which is, in Hume's opinion, typical of any passion that becomes violent (T 2.3.4.1; SBN 419).

Thirdly, the goals of Malebranche's and Hume's analyses of the passions are different. Hume aims only at explaining the passions, their causes and principles, and this explanation is not, as I said, a physiological one. Conversely, Malebranche aims not only at determining the "scope, nature [and] causes" of the passions, but also their "end, use [and] defects" (*Search after Truth*, 5.1; 340). According to Malebranche, those defects are mainly the errors to which our passions lead us regarding truth, "since judgments of passion are never in agreement with judgments of truth" (*Search after Truth*, 5.1; 340 and 5.6; 369). Here again by "the pure light of truth" Malebranche also designates "the distinct voice of our common master" (*Search after Truth*, 5.6; 369). Thus, Malebranche intends to warn against the judgments made by the passions, in order to make men "resist their seduction more easily" (*Search after Truth*, 5.6; 369, translation slightly modified). I shall insist again on this way of denouncing epistemically (and afterwards morally) the errors of the passions which is typical of Malebranche[17] but foreign to Hume. Even though Hume admits that some judgments accompanying the passions are indeed "unreasonable" (T 2.3.3.6; SBN 415–16), he does not say, as Malebranche does, that all such judgments are. More significantly, he does not make this claim at all when examining a particular phenomenon whose description he borrows from Malebranche and which I shall now examine, namely, the *self-justification of the passions*.

2. The Self-justification Principle: Hume's Borrowing from Malebranche

Whereas Malebranche devoted three chapters of *De la recherche de la vérité* to the theme of the justification of the passions, Hume chose to introduce this phenomenon in the context of a particular explanation occurring in *Treatise* 2.2.3, entitled "Difficulties solved." In the previous section, Hume had performed eight

experiments which confirmed his explanation of love and hatred by a double relation of impressions on the one hand (the relation between the *separate* pleasure or pain produced by the cause of love or hatred and the *intrinsic* pleasure or pain of those affections themselves) and ideas on the other (the relation between the *cause* of those passions and their *object*, namely, another person). Now he needs to solve some "difficulties" concerning this very explanation. Indeed, it seems that it is not simply the harm or the good that someone does us which produces our love or hatred—I do not hate someone who injures me accidentally—but only an *intentional* harm or benefit. Hume asks: does this mean that the intention of the other sensible being who is the object of our love or hatred is a supplementary and necessary cause of these passions, in addition to the double relation of impressions and ideas? He notes that "some who require not only that the pain and pleasure arise from the person, but likewise knowingly, and with a particular design and intention" add this condition (T 2.2.3.3; SBN 348).

Hume answers his question by making three points. First, the presence of an intention is *not necessary* when a constant agreeable or disagreeable quality of the other person produces our love or hatred. For example, folly and deformity are objects of our aversion but are obviously unintended. In cases involving these qualities, the double relation of impressions and ideas is sufficient to produce our love or hatred.

Secondly, when an intention *is* requisite to produce love or hatred, it is only because it reinforces this double relation of ideas and impressions which is, even in these cases, a sufficient cause of these passions. The intention thus strengthens the first and the second relation required for passions based on a double relation. Indeed, if an intentional action, agreeable or disagreeable, produces our love or hatred towards the agent while an unintended action of the same type does not, it is because an intentional action is more linked to the durable qualities of the person who is the object of these passions. Thus, because in the case of some actions the idea of the agent's intention is strongly related to the idea of the agent while the idea of the accidental and unintended action is not related to the idea of the agent, there will be some cases where the agent's intention is required. In addition, the presence of an intention also reinforces the relation of impressions, because a good done to us is all the more agreeable if it is intended, as it flatters our pride. In these cases, the intention only reinforces the double relation of ideas and impressions which gives rise to love or hatred.

Finally, experience shows that an unintended good or harm sometimes produces our love or hatred, even if these passions do not last. Here the relation of ideas between the action and the person's character is feeble, but *the absence of design does not entirely destroy this relation of ideas*, so that the relation of impressions can operate upon it.

Hume reinforces this third point with an observation that introduces the Malebranchean theme of the justification of the passions: when someone harms us, we tend to think that he is criminal, that he has bad intentions. "This," says Hume, "is a clear proof, that, independent of the opinion of iniquity, any harm or uneasiness has a natural tendency to excite our hatred, and that *afterwards we seek for reasons upon which we may justify and establish the passion.* Here the idea of injury produces not the passion, but arises from it" (T 2.2.3.9; SBN 351, my emphasis). At this point, the difficulty involved by the supposed necessity of an intention to produce our love or hatred is definitely solved. Hume concludes, "The removal of injury may remove the anger, without proving that the anger arises only from the injury" (T 2.2.3.10; SBN 351). Indeed, Hume explains that the production of the idea of injury (that is, of an *intended* harm) by the passion enables it to avoid diminution and to maintain itself, even if the passion was itself only caused by pain, that is, by *mere* harm. Moreover, it is to be noted that Hume links the theme of the justification of the passions to a more general tendency of the passions to avoid diminution, saying "Nor is it any wonder that passion shou'd produce the opinion of injury; since otherwise it must suffer a considerable diminution, *which all the passions avoid as much as possible*" (T 2.2.3.10; SBN 351, my emphasis).

The following arguments justify my claim that Hume directly borrowed this principle from Malebranche. As far as textual evidence is concerned, four points must be noted. First, Hume does not simply write that all the passions seek their own justification. More precisely, he writes that the way in which we "justify and establish the passion" consists in "seek[ing] for reasons," that is, "producing" some "ideas" or "opinions" which justify it, and which we "imagine" against all "reflection" (T 2.2.3.9; SBN 351). But this is precisely the way in which Malebranche stated this principle in *The Search after Truth*. Indeed, in the very title of Book 5, chapter 11 Malebranche announces that "all passions justify themselves in making us form judgments for their justification," he also indicates that these "false judgments" "support the imagination against the light of reason" (*Search after Truth*, 5.11; 402). Even if Hume does not precisely echo Malebranche's phrase "the passions justify themselves," stating more exactly that "we" (that is, our minds) "seek for reasons upon which we may justify . . . the passion," he also immediately adds that "passion [produces] the opinion" in question, thus echoing Malebranche's figurative way of attributing to the passions themselves the initiation of the phenomenon (T 2.2.3.9; SBN 351).

Secondly, Hume explains this principle by a cause—namely, "since otherwise [the passion] must suffer a considerable diminution, which all the passions avoid as much as possible" (T 2.2.3.10; SBN 351)—which can also be found in Malebranche. As Malebranche wrote, "[the passions] unceasingly represent to the soul the object agitating it *in the way most likely to maintain and increase its agitation*" (*Search after*

Truth, 5.10; 397, my emphasis). In short, both authors explain the self-justification principle by a non-diminution (or conservation) principle.

Thirdly, the particular passion on which Hume focuses—that is, hatred, which is excited by harm and which produces "an opinion of iniquity" to justify itself—is one of the examples used by Malebranche to illustrate his idea. Malebranche says, "we know from our own experience that we are led to form an ill opinion of those whom we do not like and to view them in a light colored by the animus of our passion" (*Search after Truth*, 5.10; 397). Even more specifically, Malebranche and Hume both take the particular example of the hatred of two antagonists in a legal contest.[18] Fourthly, another of Malebranche's examples of the self-justification principle, namely, the fact that love tends to justify itself by making us imagine that "everything about [the loved person] deserves to be loved," including his or her very "defects" and "shortcomings," which are normally an object of aversion (*Search after Truth*, 5.6; 370–71), is echoed in the *Treatise*, where Hume says "When a person is once heartily in love, the little faults and caprices of his mistress, the jealousies and quarrels, to which that commerce is so subject; however unpleasant and related to anger and hatred; are yet found to give additional force to the prevailing passion" (T 2.3.4.3; SBN 420).

There are additional, contextual reasons for thinking that Malebranche was the source of Hume's ideas about the self-justication of the passions. At least two major authors of the eighteenth century, namely Francis Hutcheson[19] and Adam Smith,[20] explicitly cite Malebranche (and no one else) as expressing the idea. Thus, in this period, the principle appears to be a Malebranchean dictum. However, precisely because Hutcheson quotes Malebranche's principle, a new possibility occurs: Hume could have found the principle in Hutcheson. Or he could have found it elsewhere, perhaps in the article on "Passion" in Chambers's *Cyclopaedia*, which includes a summary of Malebranche's theory on the passions.

These two hypotheses can both be rejected. The self-justification principle is totally absent from Chambers's article. Moreover, the two occurrences of this principle in Hutcheson's *Essay* and *Inquiry* (the second mentioning Malebranche) rephrase it in a way that Hume does not echo. Thus, in the *Enquiry*, Hutcheson maintains that to say that "all the passions and affections justify themselves" means that "while they continue, (as Malebranch [sic.] expresses it) *we generally approve our being thus affected on this Occasion*, *as an innocent Disposition*, or a just one, and condemn a Person who would be otherwise affected on the like Occasion."[21] And in the *Essay*, Hutcheson writes that "The common Effect of these *Associations* of Ideas is this, 'that they raise the Passions into an extravagant Degree, beyond the proportion of real Good in the Object: And commonly beget some secret Opinions to justify the Passions.'"[22] It is manifest that none of these versions precisely states, as both Malebranche and Hume did, that passions justify themselves *by producing opinions which justify them* and they do this *in order to avoid diminution*. Conversely,

Hume does not mention, as Hutcheson does, the approval of one's passions as innocent or the role of the associations of ideas in the justification process. In short, the way in which Hume explains the self-justification principle (all passions tend *to avoid diminution*) and elaborates on it (all passions justify themselves *by suggesting opinions which justify them*) is present in Malebranche's text and not in Hutcheson's.

Two last possibilities must still be examined. Hume could have found the self-justification principle in authors who either wrote before Malebranche or who borrowed this principle from Malebranche but (unlike Hutcheson) without citing him. As a matter of fact, I found the self-justification principle in both Arnauld and Nicole's *Logic*[23] and in Jean-Pierre de Crousaz's *A New Treatise of the Art of Thinking*,[24] one written before and one after Malebranche's *Search after Truth*. These are books that Hume had certainly (for the former)[25] or most probably (for the latter)[26] read. However, as far as textual evidence is concerned, those two versions of the self-justification principle seem less close to Hume's description of it than Malebranche's own. Arnauld and Nicole wrote about "such blindness, sophisms and illusions of the heart, which consist in transferring our passions into the objects of our passions, and in judging that they actually are what we desire that they should be"(*Logic*, 332). The self-justification principle is here obviously *in nuce*, but the idea that our passions produce opinions in order to justify and establish themselves, which we find in Hume, is still lacking. As for Crousaz's *New Treatise*, it includes both the self-justification principle and Malebranche's way of stating it. Crousaz writes, "all the passions justify themselves and by justifying them, we encourage ourselves to persevere in them, indeed by a Reason misled, but always by reason" (1:176–77). In addition, like Malebranche, Crousaz cites the examples of love and hatred (*New Treatise*, 1:178). However, unlike Malebranche, Crousaz does not explain the self-justification principle by a conservation or non-diminution principle, an important element of Malebranche's account which is echoed by Hume. It is true that Crousaz writes that "when we are given up to any passion, we suffer ourselves to see nothing but what favors that passion, and reject everything *that would obstruct or weaken it*" (*New Treatise*, 177–78, my emphasis), but this rejection of any obstruction or weakening is not attributed to the passion itself as a way to maintain its own agitation. This is a point on which Malebranche particularly insists, and which, I argue, Hume borrows from him. Finally, Crousaz does not cite the specific examples of our beloved's amiable defects and the imagined criminality of our antagonist in a legal contest. For these reasons, I conclude that Hume's borrowing from Malebranche is much more probable than his borrowing from Crousaz.

3. Hume's Appropriation of the Self-justification Principle

As I said earlier, three chapters of *De la recherche de la vérité* deal with the self-justification of the passions. In this section, I shall concentrate on some propositions

taken from these chapters in order to make a more precise comparison between Hume and Malebranche on this subject. Four differences can be discerned, three of which recall the general differences already mentioned in the first part of my paper. Naturally—even though perhaps paradoxically—those differences do not provide evidence against my previous point that Hume actually did borrow the self-justification principle from Malebranche. They rather mean that Hume appropriated this principle to fit his own non-Malebranchean interests and views on the dynamics of the mind.

First, the "clear principle" that "all the passions seek their own justification" (*Search after Truth*, 5.11; 399) is a way for Malebranche to argue that the judgments concerning good and evil which follow our passions are entirely false: such judgments, says Malebranche, "are false in every way, for the judgments formed by the passions alone are based solely on the soul's perceptions of objects in relation to itself, or rather in relation to its present emotion" (*Search after Truth*, 5.10; 397). Thus, Malebranche wanted to detect "the causes of our errors" concerning good and evil, and this is why he introduced the self-justification of the passions in the following terms: "The passions all seek their own justification; they unceasingly represent to the soul the object agitating it in the way most likely to maintain and increase its agitation" (*Search after Truth*, 5.10; 397).

As for Hume, he kept this principle of the self-justification of the passions separate from questions of truth and error, connecting the principle only with a general tendency of the passions to maintain themselves, a tendency that Malebranche had already noted in the passage just quoted. Thus, according to Hume, when the passion of hatred produces the idea of the injustice of the person who harms us, making us "imagine him criminal"—or even "bloody-minded" and "a sorcerer" (T 2.2.3.2; SBN 348)—Hume here remarks only that our passion seeks to "avoid diminution" (T 2.2.3.10; SBN 351).

A second difference concerns the way in which passions and judgments interact. Malebranche describes the way in which passions justify themselves as a process of "mutual preservation" between the passion and the judgment which causes it. He writes, "The judgment or the perception causing the passion is strengthened to the extent that the passion increases, and the passion increases to the extent that the judgment producing it is in turn strengthened. False judgments and the passions unceasingly contribute to each other's preservation" (*Search after Truth*, 5.10; 397, translation modified). This mutual preservation is due to the fact that, according to Malebranche, "judgments following and preserving the passions are exactly like those preceding and causing them" (*Search after Truth*, 5.11; 402). Such "preceding" judgments are the different judgments made about good and evil, since, as I explained earlier, Malebranche thought that all the passions have good or evil as their objects and vary according to the judgments made upon them.

Hume does not describe such a loop between the passions and the judgments that both cause and are produced by them. For Hume, the "opinion" (or Malebranche would say "the judgment") which is produced by the passion is not at the same time its original cause. For example, Hume thinks that hatred produces "the opinion of iniquity" but not that it comes from this opinion. As he explains, the cause of hatred is only the "harm or uneasiness" that naturally tends to excite our hatred (although in Hume's eyes the opinion of iniquity tends to increase the harm). This difference between Hume and Malebranche is easy to understand. As Hume had to solve some difficulties about the supposed necessity of an intention to cause love or hatred and as he wanted to defend his own explanation of these passions by a double relation of impressions and ideas, he could not at the same time adopt Malebranche's hypothesis of a "mutual conservation" of judgments and passions. Hume thus formulates a new version of the principle of the self-justification of the passions, that is, a *one-sided* production of opinions by the passions. Hume thus confirms his idea that an intention is not necessarily required to produce love or hatred: instead, he says, "the idea of injury *produces not* the passion, but arises from it" (T 2.2.3.9; SBN 351, my emphasis). Hume seems here to be consciously dissociating his view from Malebranche's, thus underlining for us the historical connection between them. However, as we shall see in the next section, he certainly has in mind another kind of loop, namely the reciprocal influences of the passions and the imagination.

A third difference between Hume and Malebranche concerning the principle of the self-justification of the passions relates to the association between different passions. Malebranche observed that passions mutually preserve themselves, writing "the passions are in no way indifferent to one another" and "all those that can exist together faithfully contribute to each other's preservation" (*Search after Truth*, 5.11; 401–2). Malebranche explained this phenomenon by citing the fact that to preserve itself each passion combines with the passions that "warmly solicit judgments in its favor" (*Search after Truth*, 5.11; 402). For example, says Malebranche, a desire for some good judged to be attainable "is animated by love, strengthened by hope, increased by joy, renewed by dread, accompanied by courage, envy, anger, and several other passions that in turn form an a infinite variety of judgments that succeed one another and sustain the desire giving birth to them" (*Search after Truth*, 5.11; 402).

Hume also remarked that passions can combine, but he explained this union by a principle of association between resembling impressions. Thus Hume writes, "All resembling impressions are connected together, and no sooner one arises than the rest immediately follow. Grief and disappointment give rise to anger, anger to envy, envy to malice, and malice to grief again, till the whole circle be compleated" (T 2.1.4.3; SBN 283). Hume observed, as did Malebranche, that "a man, who, by an injury received from another, is very much discompos'd and ruffled in his temper,

is apt to find a hundred subjects of hatred, discontent, impatience, fear, and other uneasy passions; especially, if he can discover these subjects in or near the person, who was the object of his first passion" (T 2.1.4.4; SBN 284). But Hume explained this fact by a "mutual assistance" between the association of impressions and the association of ideas (T 2.1.4.5; SBN 284). Here again, Malebranche's remarks about the association of different passions are appropriated by Hume and made to fit his own principles. In addition, Malebranche's accusation that the passions seek their own justification "against reason" (*Search after Truth*, 5.11; 402) is also absent from Hume's text. Indeed, Hume explains the association between passions in merely dynamic terms instead (that is, in terms of a "double impulse" (T 2.1.4.4; SBN 284)) or, sometimes, in chemical terms. Moreover, Hume's use of chemical terms in this context is overtly metaphorical (or, more precisely, analogical), and thus not physiological. Thus he says,

> If the objects of the contrary passions be totally different, the passions are *like* two opposite liquors in different bottles, which have no influence on each other. If the objects be intimately connected, the passions are *like* an alcali and an acid, which, being mingled, destroy each other. If the relation be more imperfect, and consist in the contradictory views of the same object, the passions are *like* oil and vinegar, which, however mingled, never perfectly unite and incorporate. (T 2.3.9.17; SBN 443; my emphasis)

Fourthly, and finally, while Malebranche developed a physiological explanation of the self-justification of the passions, Hume rejects such an explanation of this principle as well as of other principles governing the passions. Malebranche explained that if the passions can "corrupt" the mind in their favor, it is because they incessantly "agitate" the spirits which, in turn, "form deep traces" in the brain. According to Malebranche, the imagination then prevents the soul from seeing the objects as they are "in themselves," and the mind eventually adopts a judgment favorable to the passion (*Search after Truth*, 5.11; 402). Conversely, as we have seen, Hume explained the tendency of the passions to justify themselves in terms of a general tendency to preserve themselves, but he does not try to give a physiological explanation of this latter tendency.

4. The Self-justification Principle within the *Treatise*

How does the self-justification principle relate to other, more famous methodological rules or themes of the *Treatise*, and what does it add to our understanding of them? From a methodological point of view, the way in which the self-justification principle is introduced by Hume—namely, at the end of a section in which some difficulties about the particular causes of love and hatred are solved—is typical of

a more general method he adopts in the *Treatise*. This method consists in introducing a new principle when and only when it is demanded by the explanation of the phenomenon under examination (which obviously implies that already established principles fail to explain it). Indeed, to quote Hume, "to invent without scruple a new principle to every new phænomenon, instead of adapting it to the old; to overload our hypotheses with a variety of this kind; are certain proofs, that none of these principles is the just one, and that we only desire, be a number of falshoods, to cover our ignorance of the truth" (T 2.1.3.7; SBN 282).[27] Nevertheless, new principles should certainly be introduced in relevant places, that is whenever the old principles fail to explain the phenomenon under consideration. To take only one striking example, note that sympathy itself, although considered by Hume as the most remarkable quality in human nature "both in itself and in its consequences," first appears in the *Treatise* in a section on the love of fame (itself relating to a particular cause of pride—the opinions of others) and, within this section, in a preliminary circuit. The introduction of sympathy is necessary "to account for this phenomenon," that is, how the opinions of others can influence our pride (T 2.1.11.1; SBN 316). It is in the same manner that the self-justification principle appears in the *Treatise*: it is introduced in the very place where it is demanded by Hume's explanation of, or rather "solution" to, "difficulties," thus fulfilling the Newtonian prescription never to multiply causes or principles without necessity.[28] In addition, as already underlined in the preceding section of this paper, Hume's way of explaining the self-justification principle, that is, in merely dynamical terms which cautiously omit all allusion to the physiological explanation previously used by Malebranche, corresponds to another methodological maxim of the *Treatise*. It is the maxim which demands that we "rest contented with experience" (T 1.2.5.20; SBN 61; see also T intro; SBN xvii–xviii) rather than forming uncertain (though sometimes apparently "plausible") hypotheses about the way in which "animal spirits" circulate within brain traces, or how they might—as far as the self-justification principle would be concerned—cause, by their agitation, deep traces in the brain.[29]

Hume's merely dynamical description, which sometimes appears in the *Treatise* as an associationist description, should not be interpreted as less ambitious than Malebranche's own depictions of the passions which involve both a normative claim about the errors caused by our passions, which we must avoid, and a physiological account of those errors, linked as they are to the union of our minds with our bodies. For one thing, Hume's avoidance of any physiological account deeply rests on his skepticism about our capacity to account for the origin of our perceptions or the causes of their association, and Newton's maxim is a perfect illustration of this skeptical strand in the *Treatise*. Moreover, as I said in the introduction to this article, Hume's avoidance of any normative claim about the unreasonableness of our passions is deeply rooted in another aspect of his skepticism, that is,

the skeptical discovery that passions themselves are not unreasonable or contrary to reason: they are so "only so far as they are *accompany'd* with some judgment or opinion" (T 2.3.3.10; SBN 416). Besides, as we shall see presently, Hume's own ambitions regarding his explanations of the passions lie elsewhere: he aims at integrating them into a coherent and general explanation of the principles of the human mind, of which the imagination and belief are also parts.

The self-justification principle illustrates two important themes of the *Treatise*: the tendency of the mind to seek agitation and the reciprocal influences of the imagination and the passions. The self-justification principle is explained by the fact that "all passions tend to avoid diminution." Hume's recognizes that this latter principle depends on a more general tendency of the mind to seek agitation, a tendency on display elsewhere in the *Treatise*. Thus in Book 1, when explaining the transition between a present impression and a related idea which it enlivens, that is, the forming of a belief, Hume mentions an "elevation" of the spirits (which he immediately rephrases as an "enlivening of the thought") caused by the impression, a liveliness which naturally tends to communicate itself to correlated ideas (T 1.3.8.2; SBN 98–99). Moreover, Hume writes in the same section that the passions that are occasioned by those enlivened ideas also "acquire a new force and vigor" (T 1.3.8.3; SBN 99). Similarly, in section 2.2.4 of the *Treatise* which follows the section in which the self-justification principle occurs, Hume explains that the mind naturally seeks for sensations or lively ideas, especially the lively ideas of the passions of other people which, through sympathy, can become impressions. To quote Hume, the mind thus seeks for "a more sensible agitation" (T 2.2.4.4; SBN 352–53). He explains by the same principle men's "search after amusement," here echoing Pascal. It is the same idea which leads Hume to talk of sympathy as "the soul or animating principle" of all the passions when he writes, "Every pleasure languishes when enjoy'd apart from company, and every pain becomes more cruel and intolerable. Whatever other passions we may be actuated by; pride, ambition, avarice, curiosity, revenge or lust; the soul or animating principle of them all is sympathy" (T 2.2.5.15; SBN 363). To give a last example which I already mentioned, the fact that "[a] predominant passion swallows up the inferior, and converts it into itself" certainly depends on the non-diminution principle by which Hume explains the self-justification of our passions. In this case, too, Hume mentions an "agitation in the mind" by which the predominant passion grows in violence. Furthermore, Hume explicitly writes in this section that "the mind, when left to itself, immediately languishes" and that "to preserve its ardour, [it] must be every moment supported by a new flow of passions" (T 2.3.4.8; SBN 421–22). Hume thus accounts for the self-justification principle in the same way that he accounts for the forming of a belief, the search after company, the animating role of sympathy for all the passions, or the conversion of an inferior emotion into a predominant passion. In all these cases, Hume observes that the mind always tends to seek for

or to preserve whatever offers a "sensible agitation." It is true that when he evokes this agitation of the mind, Hume often mentions at the same time an *agitation of the spirits*. Nevertheless, this agitation is not taken as the physiological cause of the agitation of the mind but as a way to describe this very agitation. In all these different cases, then, Hume does not make any concession to a Cartesian physiology but simply describes the internal dynamics of the mind.[30] And the self-justification of the passions viewed as a *self-preservation* is simply part of this mental dynamics.

Secondly, the self-justification principle adds to our understanding of an important principle of the *Treatise* which is, indeed, a part of this mental dynamics, namely the reciprocal influences of the passions and the imagination. To show this, let us first recall that a section of Book 2 of the *Treatise* is entitled "Of the influence of the imagination on the passions" (T 2.3.6). Hume explains in this section that a vivid idea of good or evil not only "has an influence" on our imagination but also influences our passions, which "become more violent; and keep pace with the imagination in all its variations" (T 2.3.6.1; SBN 424). Hume follows the same line in the two following sections (T 2.3.7 and 2.3.8), in which he illustrates further this proportionality between the vivacity of our ideas of good or evil (or, in other words, "their effect on our imagination") and "their effect on our will and passions" (T 2.3.7.3; SBN 428). There he examines some more particular phenomena involving distance in time (in the past or in the future) and space, with a comparison of the effects of those two types of distances on our conceptions and passions (including special qualifications concerning admiration at T 2.3.8).[31] Here again, Hume describes this influence of distance in terms of the liveliness (of ideas) and vigor (of passions), that is, in merely dynamic terms, and not in terms of truth and error. The self-justification principle, by contrast, displays not only this "influence of the imagination on the passions" but also the converse influence. Indeed, as we have seen, this principle *first* involves the fact that our passions make us "imagine" opinions which justify them, for example, the injustice of our antagonist in a lawsuit or the malignity of the general of our enemies. Our passions thus influence our imagination.[32] *Then*, and reciprocally, the self-justification process involves a step in which this imagined opinion reinforces our passion, for instance, the opinion of iniquity gives a "new vivacity" to the idea of harm, which in its turn makes our hatred "more violent." Our imagination thus influences our passions.[33] The self-justification principle illustrates a loop (or more precisely a spiral, because a new vivacity is gained in each step) between the passions and the imagination.

The explicit description of these reciprocal influences of the passions and the imagination can be found in Book 1 of the *Treatise* in a passage perhaps more known for its description of the "first influence of the general rules" (T 1.3.13.12; SBN 150), but which also describes this no less important phenomenon:

> Let us consider the case of a man, who being hung out from a high tower in a cage of iron cannot forbear trembling, when he surveys the precipice below him, tho' he knows himself to be perfectly secure from falling, by his experience of the solidity of the iron, which supports him. . . . The circumstances of depth and descent strike so strongly upon him, that their influence cannot be destroy'd by the contrary circumstances of support and solidity, which ought to give him a perfect security. His imagination runs away with its object, and excites a passion proportion'd to it. That passion returns back upon the imagination and enlivens the idea; which lively idea has a new influence on the passion, and in its turn augments its force and violence; and *both his fancy and affections, thus mutually supporting each other*, cause the whole to have a great influence upon him. (T 1.3.13.10; SBN 148)

The self-justification principle describes a phenomenon of the same kind; to my knowledge, it is the only one in Book 2. Indeed, sections 2.3.6–8 of the *Treatise* echo it in a more incomplete way, as they only describe the influence of the imagination on our passions, whereas the self-justification principle as well as the example of the man vividly imagining his fall (which causes his fear, which, in turn, enlivens his imagination, and so on) include the whole spiral.

5. Conclusion

Unlike Malebranche's explanation of the passions, Hume's explanation is neither directed by an insistence on preserving the truth of our judgments nor based on physiology. Hume actually borrowed from Malebranche the principle of the self-justification of the passions: with Malebranche, he explains that the passions produce opinions that foster them in order to avoid diminution, and he also repeats two of Malebranche's particular examples of this principle. At the same time, Hume drops both the detection of errors and the recourse to physiology which play a large role in Malebranche's text. Instead, Hume links the self-justification principle with a mere tendency of the passions to preserve themselves by producing opinions which foster them, thus influencing our imagination which, in its turn, reinforces these passions. Hume's appropriation of the self-justification principle thus consists in viewing it, first, as an example of a general tendency of the mind to seek or preserve its agitation and, secondly, as an illustration of the reciprocal influence of the imagination and the passions, a reciprocal influence first enunciated in Book 1 of the *Treatise*, which finds its proper place in Book 2.

NOTES

I am grateful to Jacqueline Taylor, the editors of *Hume Studies* (Corliss Swain and Saul Traiger), and several anonymous referees for their helpful comments and suggestions on earlier versions of this article. A first version of this paper was presented at the 38th International Hume Society Conference in Edinburgh, in July 2011.

1 Malebranche, *De la Recherche de la vérité*, 6th ed. (Paris, 1712). This edition was present in the Physiological Library, the one Hume had access to while a student in Edinburgh: see David Hume, *A Treatise of Human Nature*, ed. David Fate Norton and Mary J. Norton (New York: Oxford University Press, 2007), vol. 2, 1019.

2 Malebranche, *De la Recherche de la vérité*, 3rd ed. (Lyon 1684). See the Norton and Norton edition of the *Treatise*, vol. 2, 1022.

3 See especially, Malebranche, *De la Recherche de la vérité*, elucidation 15.

4 Charles J. McCracken, *Malebranche and British Philosophy* (Oxford: Clarendon Press, 1983).

5 See P. J. E. Kail, "Hume's Ethical Conclusion" in *Impressions of Hume*, ed. Marina Frasca-Spada and P. J. E. Kail (Oxford: Clarendon Press, 2005), 125–39; P. J. E. Kail, "On Hume's Appropriation of Malebranche: Causation and Self," *European Journal of Philosophy* 16 (2008): 55–80; P. J. E. Kail, "Hume, Malebranche and 'Rationalism,'" *Philosophy* 83 (2008): 311–32.

6 Susan James, "Sympathy and Comparison: Two Principles of Human Nature," in Frasca-Spada and Kail, *Impressions of Hume*, 107–24. See also Peter Jones, *Hume's Sentiments: Their Ciceronian and French Context* (Edinburgh: Edinburgh University Press, 1982).

7 See the many references to Malebranche in the Norton and Norton edition of the *Treatise*, vol. 2, 1125–26.

8 Malebranche, *The Search after Truth, With Elucidations of The Search after Truth*, ed. and trans. Thomas M. Lennon and Paul J. Olscamp (Cambridge: Cambridge University Press, 1997), Preface, xxxix. Hereafter abbreviated "*Search after Truth*" and cited in text by book and chapter followed by the page number.

9 See the Nortons' annotation to section 4 of the *Treatise*'s Introduction. David Hume, *A Treatise of Human Nature*, ed. David Fate Norton and Mary J. Norton (New York: Oxford University Press, 2007), 2:691.

10 References to the *Treatise* are to David Hume, *A Treatise of Human Nature*, ed. David Fate Norton and Mary J. Norton, 2 vols. (New York: Oxford University Press, 2007), hereafter cited in text as "T" followed by Book, part, section and paragraph number; and to *A Treatise of Human Nature*, ed. L. A. Selby-Bigge, revised by P. H. Nidditch, 2nd ed. (Oxford: Clarendon Press, 1978), hereafter cited in text as "SBN" followed by page number.

11 James Fieser, "Hume's Classification of the Passions and Its Precursors," *Hume Studies* 18 (1992): 1–17. As Fieser recalls, The Stoic classification classifies the four main

passions in terms of the object (good or evil) and whether the good or evil is present or anticipated. Thus, joy is felt when a good is present, while desire is present when a good is anticipated. Grief accompanies a present evil, while fear is present when evil is anticipated.

According to Fieser, Malebranche proposes, as did Descartes and Spinoza, a revised Stoic classification. He "lists six primary passions, according to whether objects are present, anticipated, or unobtainable." These passions were joyful love (for a present good), desirous love (for an anticipated good), and sorrowful love (for an unobtainable good). Sorrowful aversion is felt with a present evil, desirous aversion, with an anticipated evil, and joyful aversion, with an unobtainable evil. The key difference between this and the traditional Stoic account is the addition of the final category of unobtainable objects" (ibid., 4). See *Search after Truth*, 5.9, 391–93.

12 See the Nortons' annotation to *Treatise* 2.1.1.3 (2:822).

13 Hume also mentions an agitation of the spirits at (T 2.2.8.4; SBN 373–74) when discussing our admiration of large objects. As far as admiration is concerned, Hume does not follow Malebranche on every point: see endnote 31 below.

14 This topic of the "agitation of the spirits" can be found for example in *The Search after Truth*, 5.3, 348, 350–51.

15 Here Hume writes "*spirits*" and not "*animal spirits*." The latter phrase can be found at T 1.2.5.20 (SBN 60), and at T 2.1.1.1 (SBN 275). In the first passage, Hume refers to some "*plausible*" physical causes of the relations of ideas or of sensations, but without generally pretending to examine such causes. In the second, the animal spirits are one among other possible causes of our sensations.

16 It is true that when Hume establishes that "every emotion, which precedes or attends a passion, is easily converted into it" (T 2.3.5.2; SBN 423), a principle which he considers as a "remarkable property of human nature" (T 2.3.4.2; SBN 419), he explains that *the excitation of the spirits* (produced by two different or contrary passions, by an external obstacle, or by uncertainty) naturally increases the predominant passion (see also T 2.3.4.2–10; SBN 419–22). But he talks indifferently of an excitation of the spirits and an "agitation *in the mind*" (T 2.3.4.7; SBN 421, my emphases), thus avoiding the physiological explanation employed by Malebranche.

17 As Jean-Christophe Bardout writes in the introduction to his French edition of Malebranche's *De la recherche de la vérité*, "warning about the sophisms dictated by the passions makes a reasonable ethics possible, first founded on knowledge, and on the love of the order of divine perfections afterwards. Thus the *Recherche de la vérité* remains the introduction to a treatise on ethics." Editor's introduction *De la recherche de la vérité*, ed. J.-C. Bardout (Paris: Vrin, 2006), 52; translated from the French.

18 Malebranche writes, "A legal contest takes place between two men to decide who owns a piece of land; they ought to produce only their titles. . . . Yet they never fail to slander one another, [and] contradict each other's statements" (*Search after Truth*, 5.6; 372). Hume writes, "Our antagonist in a law-suit, and our competitor for any office, are commonly regarded as our enemies; tho' we must acknowledge, if we wou'd but reflect a moment, that their motive is entirely as justifiable as our own" (T 2.2.3.8; SBN 351).

19 A quotation from Hutcheson's *Inquiry* that explicitly refers to Malebranche is cited below: see endnote 21.

20 "The passions, upon this account, as father Malebranche says, all justify themselves, and seem reasonable and proportioned to their objects, as long as we continue to feel them." Adam Smith, *A Theory of Moral Sentiments*, ed. D. D. Raphael and A. L. Macfie (Oxford: Oxford University Press, 1976), 157. "But our passions, as Father Malbranche [*sic*] observes, all justify themselves; that is, suggest to us opinions which justify them." Adam Smith, *History of Astronomy*, in *Essays on Philosophical Subjects*, ed. I. S. Ross (Oxford: Oxford University Press, 1980), 48.

21 Hutcheson, *An Inquiry into the Original of our Ideas of Beauty and Virtue*, ed. W. Leidhold (Indianapolis: Liberty Fund, 2004), 221, my emphasis. In other words, to quote Hutcheson's *Inquiry* again, "The Frame of our Nature on the Occasions which move these Passions, determines us to be thus affected, and to approve our Affection at least as innocent" (111).

22 Hutcheson, *An Essay on the Nature and Conduct of the Passions and Affections*, ed. A. Garrett (Indianapolis: Liberty Fund, 1982), 69.

23 Arnauld and Nicole, *Logic; or, The Art of Thinking*, part 3, xx, trans. Ozell (London, 1723).

24 J.-P. de Crousaz, *A New Treatise of the Art of Thinking* (London, 1724).

25 Arnauld and Nicole's *L'art de penser* is cited by Hume in T 1.2.4.12n10 (SBN 43).

26 See the Nortons' annotations referring to Crousaz's *New Logic*: ann. 4.3–6, 26.39–40, 51.42, n. 20.3–4 (67), 262.12–13, 262.37–38, 286.38–9.

27 This principle of parsimony also reads: "It [is] an inviolable maxim in philosophy, that where any particular cause is sufficient for an effect, we ought to rest satisfy'd with it, and ought not to multiply causes without necessity" (T 3.3.1.10; SBN 578).

28 Newton is named by Hume on this matter in *An Enquiry Concerning the Principles of Morals*: "It is entirely agreeable to the rules of philosophy, and even of common reason; where any principle has been found to have a great force and energy in one instance, to ascribe to it a like energy in similar instances. This is indeed NEWTON's chief rule of philosophizing. [*Princip*., lib. 3]." David Hume, *An Enquiry Concerning the Principles of Morals*, ed. Tom L. Beauchamp (New York: Oxford University Press, 1998), section 3, paragraph 48.

29 The explanation which occurs at T 1.2.5.20 in which Hume refers to the animal spirits and the traces of the brain thus remains, as Hume himself admitted, "an exception" to this maxim.

30 See the very last words of Hume's *Dissertation on the Passions*: "It is sufficient for my purpose, if I have made it appear, that, in the production and conduct of the passions, there is a certain regular mechanism, which is susceptible of as accurate a disquisition, as the laws of motion, optics, hydrostatics, or any part of natural philosophy." *A Dissertation on the Passions*, 6.19, ed. T. L. Beauchamp (Oxford: Clarendon Press, 2007), 29. In *Dissertation* 2.47, Hume speaks of an "internal mechanism."

31 Compare, for example, Malebranche's observations on admiration in *Search after Truth*, 5.7, where he insists on the errors occasioned by the idea we have of the greatness or smallness of ourselves or other objects, contrary to the "pure idea of truth" (5.7; 381–82) with Hume's neutral and plainly descriptive treatment of the same phenomenon in T 2.3.8.2 (SBN 432–33).

32 In Book 1 of the *Treatise*, Hume has enunciated a principle which depends on this influence of the passions on the imagination, namely their influence on *belief*: See T 1.3.10.4 (SBN 120).

33 Note that those two steps do not involve, as Malebranche said, that "judgments following and preserving the passions are *exactly like those* preceding and causing them" (*Search after Truth*, 5.11; 402). As already noticed, using the example of hatred, Hume says, on the contrary, that "*the idea of injury produces not the passion, but arises from it.*" See section 2 above.