



Kant's Argument for the Apperception Principle

Melissa McBay Merritt

Abstract: My aim is to reconstruct Kant's argument for the principle of the synthetic unity of apperception. I reconstruct Kant's argument in stages, first showing why thinking should be conceived as an activity of synthesis (as opposed to attention), and then showing why the unity or coherence of a subject's representations should depend upon an *a priori* synthesis. The guiding thread of my account is Kant's conception of enlightenment: as I suggest, the philosophy of mind advanced in the Deduction belongs to an enlightenment epistemology. Kant's conception of enlightenment turns on the requirement that a subject be able to recognize herself as the *source* of her cognitions. The argument for the apperception principle is reconstructed under the guidance of this conception of the ideal of enlightenment.

If by the word 'understanding' is meant the faculty of cognition by means of rules [...] then these rules are not to be understood as those according to which nature guides the human being in his conduct, as occurs with animals which are driven by natural instinct, but only those that he himself *makes*. (Kant, *Anth* §42 (7:197))

1. Introduction

The fundamental principle of Kant's epistemology is called the 'principle of the original *synthetic* unity of apperception' (§17, B137). I shall simply refer to it as the 'apperception principle'.¹ Kant maintains that it is the 'highest' and 'first principle' of human cognition (§16, B135; §17, B139). Although Kant does not offer a concise statement of this principle, let me begin with what I take to be an uncontroversial report of its content: the apperception principle says that a cognitive subject's representations belong to it in virtue of an *a priori* synthesis.²

What is crucial here, as some commentators have noted, is that the apperception principle makes a claim about 'a priori synthesis'. One reason why this is puzzling is that we generally speak of propositions or modes of knowledge as being *a priori*—independent of experience—so it is unclear what it

European Journal of Philosophy ■■■■ ISSN 0966-8373 pp. 1–26 © 2009 Blackwell Publishing Ltd., 9600 Garsington Road, Oxford OX4 2DQ, UK and 350 Main Street, Malden, MA 02148, USA.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000

could mean for synthesis to be a priori. At the end of the paper I will address Kant's claims about a priori synthesis in particular. The bulk of the paper, however, will be devoted to considering the significance of Kant's appeal to synthesis *as such*—an issue that is not often directly addressed by commentators, and which we need to come to terms with before considering the significance of calling the synthesis in question 'a priori'. One of the major moves Kant makes in the Transcendental Deduction is to suggest that thinking must be conceived as an activity of *synthesis*—a point which is encapsulated in the apperception principle. As I will argue, this conception of the activity of thinking is advanced in opposition to a conception of thinking as *attention*. The attention model of thinking, I will suggest, is rejected on the grounds that it is not compatible with the enlightenment aspirations of a properly 'critical' philosophy. Much of what I shall say about Kant's apperception principle will stem from this point.

Let me say outright that I do not find an explicitly delineated argument for the apperception principle in the pages of the *Critique*—though my aim here is to reconstruct one from materials that Kant provides.³ Commentaries often present the apperception principle as an unargued, or self-evident, presupposition of Kant's Transcendental Deduction.⁴ We should be wary of this suggestion: it should hardly be appropriate for Kant to present the fundamental principle of his epistemology (and, by extension, theoretical metaphysics) as an unargued presupposition—the principle simply does too much work in Kant's system to be granted such easy passage. That being said, it is the case that the Deduction kicks off with a putatively self-evident claim about self-consciousness; and since Kant takes apperception to be a mode of self-consciousness, it is true that the argument of the Deduction proceeds from a claim about apperception.⁵ Yet that initial claim is not identical with the apperception *principle*, as we shall see: the starting point of the Deduction contains no claim about a priori synthesis. Taking the apperception principle to be an unargued presupposition confuses claims that Kant keeps distinct. Part of what I shall be doing in this paper is to pry apart and explore the argumentative distance between that starting point and the apperception principle itself.

My reconstruction of Kant's argument for the apperception principle rests on two related presuppositions about the 'critical philosophy' to which it belongs. First, the *Critique of Pure Reason* is billed as human reason's assessment of its own cognitive capacity. The genitive construction in the title (*Kritik der reinen Vernunft*) is ambiguous, suggesting that reason may be both the subject and object of the inquiry. This point will inform my interpretation of Kant's use of the first-person in the argument for the apperception principle.⁶ The second point can also be drawn out of the title, now focusing on the term 'critique': at issue is Kant's conception of a properly *critical* philosophy. By a critical philosophy in this generic sense, I mean a philosophical project that properly acknowledges the demands of enlightenment. For Kant, as for many others, the motto of enlightenment is to 'think for oneself'—and not accept claims passively on the basis of external authority. This ideal is invoked in various ways throughout the *Critique*: in its initial portrayal of the scientist as a judge putting nature to the

question, rather than as a pupil passively registering information from without (Bxiii); in its suggestion that an individual cannot be said to possess rules or principles for knowledge unless she is able to recognize, independently, their applicability in various contexts (A134/B172-5); and in its spoof of the student of dogmatic metaphysics, who is able to make sanctioned moves within a given system but is dumbstruck should someone 'dispute a definition with him' (A836/B864). Yet the ideal of enlightenment is writ most large in Kant's conception of critical philosophy as an exercise of corrective self-knowledge—an exercise carried out under the motto that only human reason can adjudicate the claims of human reason (A752/B780; A849/B877; *P* 4:263). Now, if Kant takes the apperception principle to be the fundamental principle of human knowledge, then presumably the account of our cognitive capacity that arises from this principle should be appropriate to the enlightenment ideal. In other words: the apperception principle belongs to an enlightenment epistemology.

As I will suggest, Kant's conception of enlightenment provides the general framework within which he argues for a particular view of the nature of thinking—the view, indeed, that is expressed in the apperception principle. I begin by sketching that framework, focusing on a set of three maxims that specify fundamental normative requirements of our cognitive practices (§2). An enlightenment epistemology, as Kant understands it, gives special emphasis to the notion of cognitive agency—a point that shall prove important in the argument for the apperception principle. I then turn to the text of the Deduction: at first my aim is simply to draw attention to the argumentative distance between the starting point of the Deduction and the apperception principle itself (§3). Then I prepare to reconstruct the argument for the apperception principle, first by considering the two models of thinking that I take to be at issue for Kant: thinking as a certain mode of attention, and thinking as an activity of synthesis. After providing an initial account of why the attention model of thinking should be rejected (§4), and clarifying crucial terminology (§5), I reconstruct Kant's argument for the apperception principle. This reconstruction is broken down into two main parts: first I explain why Kant supposes that thinking must be conceived as an activity of synthesis (§6), and then I address the crucial claim about *a priori* synthesis (§§7–8).

2. The Prospect of an Enlightenment Epistemology

Kant's conception of enlightenment is best known through his popular essay, 'What is Enlightenment?'. There Kant presents the idea of enlightenment in terms of a command: 'Have courage to make use of your *own* understanding!' (8:35). The motto expresses the familiar view that an enlightened mind is one that asserts its own power of thought, rather than succumbing to unreflective prejudice of various kinds. This idea is reformulated in terms of an opposition between mechanism and spontaneity: when we take things to be a certain way based on prejudice—the sources of which are named as *imitation*, *custom*, and *inclination* (JL 9:76)—our view of things is determined by an automatic, stimulus-

1 driven response of some kind. We suppose that something is worthy of imitation
 2 because it is prestigious, we take a certain course of events to be necessary
 3 because that is how it has always come to pass, and we suppose that something is
 4 good because of our habitual desire for it.⁷ When we take things to be a certain
 5 way on the basis of prejudice, we fail to be reflective: for Kant, this means that we
 6 fail to recognize that we are taking things to be a certain way 'based on subjective
 7 causes' rather than 'objective grounds' (9:76). The recognition of how our
 8 thinking was overtaken by the mechanical operation of prejudice provides the
 9 opportunity to reassert mastery over our own minds. Thus the call to think for
 10 oneself can be interpreted as the requirement that we exercise the spontaneity or
 11 freedom that is proper to a rational mind.⁸

12 However, enlightenment is not tantamount to unbridled freethinking. An
 13 enlightened individual is not one who simply thinks for herself, willfully
 14 disregarding even her own experience and heedless of the claims and testimony
 15 of others. For this reason, while Kant trumpets the importance of thinking for
 16 oneself at the outset of the essay on enlightenment, the bulk of the essay concerns
 17 the importance of communicating one's thoughts in public discourse. Our
 18 readiness to think for ourselves must be matched by our readiness to consider
 19 other cognitive claims. Still, Kant's account of enlightenment in the famous essay
 20 is only partially complete. For in a variety of other texts, Kant expresses the ideal
 21 of enlightenment in terms of a battery of *three* maxims of 'common human
 22 understanding' that specify the fundamental normative requirements of our
 23 cognitive practices.⁹ The famous essay on enlightenment implicitly addresses
 24 only the first two: *to think for oneself* and *to think in the position of everyone else*.¹⁰

25 The third maxim is *to think always consistently with oneself*. Kant takes it to be a
 26 synthesis of the first two (5:295): it draws out the implications of what it is to
 27 think for oneself in the sphere of public discourse. It is constitutive of the
 28 enlightenment conception of thinking that we *strive for*, and to some extent
 29 *actually achieve*, unity and coherence in light of both our own experience and the
 30 claims of other judging subjects. Inconsistent representations can be maintained
 31 by a subject only to the extent that she is either in the dark about their
 32 inconsistency or about their being *her own* representations. Thus Kant presents
 33 the three maxims as a package, and together they articulate the enlightenment
 34 ideal that the subject recognize herself as the *source* of her cognitions. In turn, the
 35 possibility of being held responsible for one's cognitive claims—a basic
 36 requirement of reasoned public discourse—rests on this reflective self-awareness.

37 Together the maxims point to a normative ideal, the flourishing of human
 38 understanding. They also effectively concern the *nature* of the human under-
 39 standing, inasmuch as a subject with sound understanding will at least implicitly
 40 recognize the maxims as binding on her cognitive practices. Since the maxims
 41 concern how things ought to be with our cognitive practices, then any account of
 42 the activity of thinking that is put forward in the Transcendental Deduction should
 43 be adequate to the enlightenment ideal that is embodied in the three maxims.

44 Later on, I aim to show how Kant's account of the activity of thinking—the
 45 account that is bound up in the apperception principle—is guided in important

ways by his conception of the ideal of enlightenment. One point in particular will prove crucial. Together, the three maxims come down to a point about self-consciousness: namely, the normative requirement that a subject be able to recognize herself as the *source* of her own thought. This point has much to do with the notion of cognitive agency, which will figure in my reconstruction of Kant's argument for the apperception principle in the final sections of the paper. Let us turn now to the text of the Deduction so that we may gain an initial view of what Kant is arguing for.

3. The Cogito Statement and the Apperception Principle Distinguished

The Deduction begins with a claim about self-consciousness, uttered in the first person: 'The *I think* must be *able* to accompany all of my representations' (B131).¹¹ This remark is often mistaken for the apperception principle. To prevent our confusing what Kant distinguishes, I introduce the following terminology:

- (a) **cogito statement** shall refer to the famous line, 'The *I think* must be *able* to accompany all of my representations';¹² while
- (b) **apperception principle** shall refer to the principle of the synthetic unity of apperception. The apperception principle claims that all of a single subject's representations belong to it in virtue of an a priori synthesis.

Let us begin with the cogito statement. It is a claim, in the first person singular, regarding something common to all of the subject's representations. The commonality is the 'I think', which must be able to accompany each one. The cogito statement is established through a partially suppressed reductio. We are led to consider the alternative scenario in which the 'I think' *need not* be able to accompany all of 'my' representations. That Kant is drawing out the implications of the denial of the cogito statement becomes clear in the remark that follows: 'For otherwise something would be represented in me which could not be thought at all, which is as much as to say that the representation would either be impossible, or at least nothing for me' (B131-2). Our acceptance of the cogito statement rests upon the presupposition that the 'I' that speaks is an intellect: a representation that it could 'have' without possibly recognizing itself as 'thinking' could not be recognized by it *as* a representation at all.¹³ The use of the first person, I take it, reminds the reader that the general project of the *Critique* is one of self-knowledge: the 'I' that speaks is an intellect assessing its own cognitive capacity. In this context, the cogito statement figures as a self-evident proposition.¹⁴

It is important to note that the cogito statement is not equivalent to the apperception principle, because it makes no claim about synthesis. In order to appreciate how the apperception principle comes on the scene (though not yet how it is argued for), we must acknowledge the specific investigative framework of the *Critique*: it is pure reason's assessment of its capacity for *theoretical*

cognition. Since such cognition concerns objects that ‘must be given from elsewhere’ (Bx), such objects must be represented *as* given if they are to be known at all. That is, they must be represented in a manner that respects their independence (as far as their existence is concerned) from this cognition. Hence the theoretical capacity of pure reason involves two elements: reason must not only include the intellectual element that is naturally associated with it, but it must also include a faculty of sensibility. Given that the *Critique* is a project of self-knowledge—pure theoretical reason assessing its own cognitive capacity—the cogito statement can be considered as uttered by pure theoretical reason. Since the cogito statement pertains to all of this subject’s representations, it would therefore pertain to some subset of them as well. In this way, the cogito statement can then be brought to bear on *sensible* representation in particular: this move brings the apperception principle into view.

Noting that sensible intuition is ‘[t]hat representation which can be given before all thinking’ the move is made: ‘Thus, everything manifold in intuition has a necessary relation to the *I think* in the same subject in which this manifold is encountered’ (B132). Therefore, whatever may figure as sensible representation is always already subject to the conditions of its being thought. So, Kant continues: ‘as my representations (even if I am not conscious of them as such) they must necessarily be adequate to the condition under which they *can* stand together in a universal self-consciousness, for otherwise they would not throughout belong to me’ (B132). From here, Kant claims that representations belong to a single subject in virtue of a synthesis that he calls ‘original combination’ (B133).

In overview: the cogito statement registers as a claim uttered by pure theoretical reason, the subject of the critical self-examination. This intellect then considers how it is possible for it to enjoy sensible representations. The answer to its question is the apperception principle: sensible representations belong to it in virtue of a necessary, a priori synthesis. Thus the apperception principle comes on the scene when the cogito statement is brought to bear on sensible representations in particular.

Let me stress that the above is not an argument for the apperception principle; it is merely a presentation of its place in the order of exposition. It allows us to appreciate the difference between the cogito statement and the apperception principle in the following terms. The cogito statement does not specify the conception of thinking in play. The cogito statement would hold whether thinking is conceived as an act of attention, or as an act of synthesis—which I take to be the two competing conceptions of thinking at issue for Kant.¹⁵ The apperception principle, by contrast, invokes a specific conception of thinking: thinking as an activity of *synthesis*—indeed, as an activity involving *a priori* synthesis.

4. The Attention Model of Thinking

In the previous section, I drew attention to the difference between the starting point of the Deduction—the cogito statement—and the apperception principle itself. Only with the apperception principle does a claim about synthesis arise. In

the rational subject—the agent of the critical investigation—without the resources to account for the objective content of its representations.²³

This brings us, then, to the Kantian alternative: thinking is an activity of synthesis, or *combining* representations. The question we face now is whether the synthesis model of thinking holds out any special promise for meeting that ‘critical’ standard Kant sets for himself when he says that only human reason can adjudicate the claims of human reason. Does the synthesis model of thinking provide the resources to *account for*, rather than assume, the contentfulness of our representations? For the synthesis model to hold out such promise, it must not follow the attention model in having the content of representations be given independently of the activity of thinking itself. That would mean that the content of representations would in some sense be constituted by the very activity of thinking.

In §6, I will provide a positive argument for the synthesis model of thinking. So far, I have only provided an initial explanation of why the attention model of thinking ought to be rejected. I have done so without entering into the details of the Deduction—drawing, rather, only on an abstract consideration of this conception of thinking, and showing its general unsuitability for an enlightenment epistemology. In order to appreciate the positive argument for the synthesis model of thinking—the first step in my reconstruction of the argument for the apperception principle—it will be necessary to consider the details of the opening stretch of the Deduction. At that point, I will take up the guiding thread of my interpretation once again: namely that the philosophy of mind advanced in the Deduction belongs to an enlightenment epistemology, and thus must acknowledge the ideal of enlightenment. As I suggested in §2, that ideal most fundamentally concerns the subject’s capacity to recognize herself as the *source* of her representations. In the Deduction, this issue comes into play in the context of Kant’s grappling with the problem of personal identity. There Kant seems to advocate for the apperception principle on the grounds of certain considerations about personal identity. By showing how the issue of personal identity bears on the enlightenment ideal in question, I will show that Kant has a positive argument for the synthesis model of thinking—and, in the end, for the apperception principle itself.

Before turning to that task, it will be necessary to examine Kant’s usage of crucial terminology—most notably *thinking*, *cognition*, and *consciousness*.

5. Clarification of Terminology

Let us begin with the notion of thinking, since it appears prominently at the outset of the Deduction, in the cogito statement. In the *Jäsche Logic*, Kant presents thinking in tandem with the understanding, suggesting that we come to terms with the two terms concurrently.

Like all our powers, *the understanding* in particular is bound in its actions to rules, which we can investigate. Indeed, the understanding is to be regarded in general as the source and the faculty for thinking rules in

1 general. For as sensibility is the faculty of intuitions, so the under-
 2 standing is the faculty for thinking, i.e. for bringing the representations of
 3 the senses under rules. (9:11)
 4

5 In this passage, Kant specifies the activity of thinking as 'bringing the
 6 representations of the senses under rules'. This may seem to be an overly narrow
 7 gloss on thinking, since presumably we can think of notional things, without
 8 having sensible representations in play. In the *Critique's* Preface, Kant remarks:
 9 'I can *think* whatever I like, as long as I do not contradict myself'—whereas in
 10 cognition my thought must be valid of objects (Bxxvi).²⁴ With this in mind, I will
 11 suggest that we make Kant's gloss from the *Logic* a bit more general: thinking is
 12 simply the activity of representing according to rules, whereas cognition is
 13 thought that is valid of objects.

14 In claiming that the understanding 'is bound in its actions to rules' Kant's point
 15 is not that the understanding is *distinguished* by the fact that it operates according
 16 to rules. For the *Logic* begins with the stark claim that everything in nature—and
 17 indeed everything that we can come to cognitive terms with—operates according
 18 to rules. So his point is not that the understanding operates according to rules,
 19 whereas our capacity for sensible representation somehow does not. Rather, his
 20 point is that anything that we can come to cognitive terms with we represent
 21 according to rules, and the understanding is the *source* of such rules.²⁵

22 Kant seems also to suppose that the capacity to represent things according to
 23 rules entails a capacity to represent the rules themselves: for he says that the
 24 understanding is 'the faculty for thinking rules in general', suggesting that the
 25 rules can themselves be brought to consciousness. Kant sheds some light on this
 26 point later in the *Logic*, in a passage that presents seven 'degrees' by which a
 27 representation can relate to objects. The first three stages concern modes of
 28 representation that Kant supposes can also be attributed to animals:
 29

30 In regard to the objective content of our cognition in general, we may
 31 think the following *degrees*, in accordance with which cognition can, in
 32 this respect, be graded:

33 The *first* degree of cognition is: *to represent* something;

34 The *second*: to represent something with consciousness or *to perceive*
 35 (*percipere*);

36 The *third*: *to be acquainted* with something [etwas *kennen*] (*noscere*), or to
 37 represent something in comparison with other things, both as to
 38 *sameness* and as to *difference*[.] (JL 9:64–5)

40 Since the notion of consciousness is introduced only in the second level, we must
 41 first consider what it would be to represent something without consciousness.²⁶
 42 An example of this might be what we represent in dreams that we never recall;
 43 another example might be Leibnizian *petites perceptions*—e.g. that we represent
 44 the sound of each wave as we stand on the shore, even though we are only
 45 conscious of their aggregated pounding sound.²⁷

1 But how, exactly, does Kant understand the notion of consciousness? Kant's
 2 gloss in the *Logic* is not altogether helpful: 'Consciousness is really a
 3 representation that another representation is in me' (9:33). One problem with
 4 this gloss is that it presupposes the capacity for first-personal representation,
 5 which Kant attributes only to rational beings.²⁸ Since Kant evidently attributes
 6 conscious representation to non-rational beings, there must be some way to
 7 render the notion of consciousness that does not presuppose self-consciousness
 8 and first-personal representation. A clue is made available in the *Anthropology*,
 9 where Kant notes that a child, in the normal course of his development, first
 10 refers to himself in the third person, and only later in the first person: 'Before he
 11 merely *felt* himself, now he *thinks* himself' (7:127). The determination of self that
 12 is merely felt and not thought would not be represented in terms of agency;
 13 presumably, it would be represented merely as a locus of affection. If so, animal
 14 representation with consciousness (i.e. the second level in Kant's list) is
 15 something's figuring in some such locus of affection. The last level of
 16 representation that could still belong to an animal mind is the third: this is 'to
 17 be acquainted with something' (*etwas kennen*), and it involves representing
 18 something in its sameness to and difference from other things.²⁹

19 Now, let us consider how the progression continues, to see how Kant
 20 distinguishes the consciousness proper to a rational mind from the consciousness
 21 that can belong to an animal mind:
 22

23 The *fourth*: to be acquainted with something *with consciousness*, i.e., to
 24 *cognize* it (*cognoscere*). Animals are acquainted with objects too, but they
 25 do not *cognize* them.

26 The *fifth*: to *understand* something (*intelligere*), i.e., to cognize something
 27 *through the understanding by means of concepts*, or to *conceive*. One can
 28 conceive much, although one cannot comprehend it, e.g., a *perpetuum*
 29 *mobile*, whose impossibility is shown in mechanics.

30 The *sixth*: to cognize something through reason, or to *have insight* into it
 31 (*perspicere*). With few things do we get this far, and our cognitions become
 32 fewer and fewer in number the more that we seek to perfect them as to
 33 content.

34 The *seventh, finally*: to *comprehend* something (*comprehendere*), i.e., to
 35 cognize something through reason or *a priori* to the degree that is
 36 sufficient for our purpose. (9:65)

37 Now, if merely being acquainted with something is to distinguish it from others
 38 according to sameness and difference, then presumably being acquainted with
 39 something *with consciousness* would be to represent the distinguishing itself. Kant
 40 calls this *cognition*: the representation of an object as distinct from other objects,
 41 where the subject is conscious of the distinguishing itself, though not yet (or not
 42 necessarily) of the rule operative in the distinction. In other words, the subject
 43 recognizes that she distinguishes an object, but the distinguishing is not itself
 44 understood. Accordingly, the next level is *understanding*, where the subject does

not merely represent that she distinguishes one thing from another, but also has at least a tacit grasp of the rule determining the representation. But mere understanding itself falls short of rational cognition, which is at issue in the final two stages. As Kant takes reason to be a faculty of inference, presumably we have 'insight' into something when we not merely determine it under a rule, but consider it and the rule under which it is determined in relation to a broader inferential whole. In this inferential context, the rule by which the object is determined can itself be understood. Finally, we comprehend something when our determination of the object figures in a system of knowledge that is based upon a priori principles.³⁰

One important lesson of the *Logic* passage is that the term 'consciousness' in Kant's usage must be specified with regard to whether the representation in question is an actualization of either animal or broadly rational capacities. By the latter I mean to refer to the representational capacities distinctive of a rational being: i.e. the capacities included in Kant's conception of the 'higher cognitive faculty' or the intellect.³¹ Rational representation involves the possibility of bringing *rules* of representation themselves to consciousness. Moreover, the rational mind's consciousness of a rule is not an isolated affair: rules that are brought to consciousness are considered in light of an idea of a whole of cognition. This, at any rate, is *comprehension*, the final 'grade' in Kant's list.

Now, Kant says that everything in nature, and indeed everything that we can come to cognitive terms with, operates according to rules. Thus, animal representation operates according to rules. But it does not operate according to rules that the animal subject makes: hence animal representation does not incorporate the possibility of representing the rules themselves. This means, I take it, that animal consciousness could be understood according to the 'attention' model considered above.³² Mental content can only figure as *given* for such a creature: content is something thrust before it, to which it may respond; but content is not something that it constitutes itself, and so is able to assess. By contrast, rational consciousness involves the possibility of assessing the rules of cognition themselves—and thereby the content of thought—in light of some idea of a coherent and unified whole of cognition. Now, if the synthesis model of thinking is one according to which content arises from the synthetic activity of thought—i.e. that representations are *united to a certain content*—then the possibility of rational consciousness, as it is explained in this passage from the *Logic*, would seem to require the synthesis model of thinking. With this in mind, let us return to the Deduction, and begin to reconstruct Kant's argument for the apperception principle.

6. Personal Identity and the Synthesis Model of Thinking

My general aim, as noted at the outset, is to reconstruct Kant's argument for the apperception principle. To make that task more manageable, I propose to take the apperception principle in stages. The apperception principle contains the following three claims:

- (1) thinking is to be conceived as an activity of synthesis;
- (2) the general coherence or unity of a subject's representations is 'synthetic', or arises from the synthetic activity of thought;
- (3) this unity is made possible by *a priori* synthesis.



We have already seen why the attention model of thinking is rejected: it does not cohere with the enlightenment aspirations of the critical project. The rejection of the attention model allows us to appreciate more clearly the move from the cogito statement to the apperception principle sketched in §3 above. Nothing in the cogito statement entails anything about synthesis: by the lights of the cogito statement alone, representations could belong to the subject simply in virtue of an act of *attention* possibly accompanying each one. The recognition that the attention model does not accord with the enlightenment aspirations of the critical project supports the specification that is made in moving from the cogito statement to the apperception principle.

But we still need a conclusive argument that thinking *must be* conceived as an activity of synthesis: the mere rejection of the attention model of thinking, on the grounds that it does not cohere with the enlightenment ideal that informs Kant's critical project, does not force the synthesis model of thinking upon us. As it turns out, the requirement that we conceive of thinking as an activity of synthesis cannot be neatly separated from the second point contained in the apperception principle, regarding the 'synthetic unity' of the subject's representations. In this section, I shall address both issues together. This will lead us into further complications of Kant's account of the apperception principle—most notably the issue of personal identity. As I will argue here, this is neither an accident nor a gratuitous change of topic. Indeed, if we read Kant's remarks about personal identity in the context of his preoccupation with the ideal of enlightenment, we find in them the resources for a positive argument for the synthesis model of thinking.

Earlier, in §4, I presented an argument against the attention model of thinking without drawing upon the details of the Transcendental Deduction: I argued that the attention model of thinking cannot contribute to an enlightenment epistemology, at least on Kant's understanding of such a project. Here I aim to show how that initial argument is bears upon the stretch of the Deduction in which Kant aims to establish the apperception principle. As we saw, Kant expresses the ideal of enlightenment through the three maxims of common human understanding. Together, the three maxims come down to a point about self-consciousness: namely, the normative requirement that a subject be able to recognize herself as the *source* of her own thought. On the face of it, this requirement might not seem to have much to do with the issue of personal identity, the topic that occupies Kant throughout the stretch of the Deduction in which he tries to establish the apperception principle. However, as I will suggest, Kant's discussion of personal identity in the Deduction concerns precisely the conditions under which a subject can recognize herself as the source of her own representations.

The gist of Kant's argument is this: if the subject's enjoyment of representations is understood according to the attention model of thinking, then it does not follow from a subject's enjoying a totality of representations that she can recognize herself as the *source* of these representations—because it does not even follow that she can recognize herself as the *unitary subject* of these representations throughout. Hence the following passage, from §16 of the Deduction:

For the empirical consciousness which accompanies various representations is in itself dispersed and without relation to the identity of the subject. Such relation does not yet come about by my accompanying each representation with consciousness, but rather by my *adding* [*hinzusetze*] and being conscious of the synthesis of them. (B133)

If representations belong to the subject merely in virtue of an act of attention possibly accompanying each one, then instead of a unitary subject, we might as well have an aggregate of representations. If we adopt the attention model of thinking, then we leave open the possibility of a merely 'aggregated'—rather than unified—subject. And if our epistemology draws upon such a philosophy of mind, then this epistemology could not, in turn, account for the requirement that is embodied in the enlightenment ideal: namely, that a subject be able to recognize herself as the *source* of her representations. If this is correct, then an enlightenment epistemology requires a different conception of thinking.

With this in mind, let us consider Kant's initial claim about *cognition* from the first-edition Deduction:

If every individual representation were completely foreign to every other, as it were, isolated and separated from every other, then there would never arise anything like cognition, which is a whole of compared and connected representations. (A97)

The passage introduces a minimal conception of cognition—one that even a thoroughgoing empiricist could endorse—and points out what it entails. Kant claims that cognition is a *whole*: a unity comprised of representations that minimally *bear comparison* to one another. Such a 'unity' of representations could be distinguished from an accidental aggregate: a collection of representations that do not necessarily bear comparison to one another. Now, the attention model of thinking allows for the possibility of such accidental aggregates: it allows for the possibility of representations that are 'completely foreign' to, and 'isolated and separated' from, each other. The representations in question might not bear comparison to each other at all, but may only belong together in virtue of a subject's mere act of attention possibly accompanying each one.

Some such possibility seems to be envisaged by Hume, in the skeptical phase of his account of personal identity in the *Treatise*.³³ Famously, Hume refers to the mind as 'nothing but a heap [. . .] of different perceptions' (1.4.2.39) and later to the self as 'nothing but a bundle [. . .] of different perceptions, which succeed each other with an inconceivable rapidity, and are in perpetual flux and movement'

(1.4.6.4).³⁴ Invoking heaps and bundles suggests that the mind is to be conceived as an aggregate of contents. The attention model of thinking accords with such a conception of the self because it does not require that the subject's representations even minimally bear comparison to one another.

For our enjoyment of representations to have robust implications about the identity of the subject, the synthesis model would be required. On the synthesis model of thinking, the content of any particular representation depends upon a synthetic unity of representations. Thus, the enjoyment of any contentful representation entails an activity of synthesis on the part of the subject, in virtue of which that representation has its content as part of a system. As Kant remarks, the synthetic activity of thought allows for 'the *identity of consciousness*' to be represented '*in these representations themselves* [. . .]' (B133). In the enjoyment of any representation, I am necessarily able to represent myself as the synthesizer of the totality of which it is part. I take this to be the point of the next stretch of §16:

The thought that these representations given in intuition belong one and all to me means accordingly the same as that I unite them in one self-consciousness, or at least can unite them in it, and although it is itself not yet the consciousness of the *synthesis* of representations, still it presupposes the possibility of the latter, i.e., only because I can grasp the manifold of them in one consciousness do I call them one and all *my* representations; for otherwise I would have as multi-colored, various a self as I have representations of which I am conscious. (B134)

Again we can see how the attention model of thinking clashes with the enlightenment ideal. For the attention model of thinking allows for the possibility of accidental aggregates of representations. The elements of the 'multi-colored' collection need not have anything in common except for the possibility of the 'I think' accompanying each one. The attention model of thinking is indifferent to disunity in the subject: the subject may well be the 'multi-colored, various' self that Kant refers to here.

Thus we can recast the argument against the attention model of thinking. The attention model allows for the skeptical non-account of personal identity that Hume offers with his 'heap' and 'bundle' metaphors. Since, by the enlightenment ideal, a subject must be able to recognize herself as the *source* of her representations—and hence as a unitary subject—a model of thinking that is indifferent to disunity in the subject cannot be endorsed. We are now in a position to advance a positive argument for the synthesis model of thinking. To take thinking to be an activity of synthesis is to suppose that the coherence of representations is *wrought* by the subject in the very activity of thinking. And so, if the issue of the coherence of the cognitive subject's representations is fundamental to the account of cognition, and if the synthesis model of thinking allows us to account for this necessary coherence while the attention model fails, then that should be a decisive score for synthesis.

Of course, the apperception principle says considerably more than simply that thinking must be conceived as an activity of synthesis: through it, Kant claims that the unity of a cognitive subject's representations depends upon an *a priori* synthesis. It is time to take up that issue. However, I shall do so in a somewhat unorthodox way, since I shall dwell a bit longer with Hume. As I will suggest, Hume comes rather close to the apperception principle—even though he would never be able to accept any claim about a priori synthesis. By examining Hume's account a bit further, I hope to reveal the extent to which Kant's apperception principle rests on a relatively uncontroversial conception of the ideal of enlightenment. Appreciating this point will put us in position to clearly ascertain both the meaning, and the basis, of Kant's appeal to a priori synthesis by the end.

7. The Enlightenment Ideal and Cognitive Agency: Ambiguities in Hume's Account

As I have just suggested, the attention model of thinking is indifferent to disunity in the subject. Thus, if we adopt the attention model of thinking, we have no ready resources to resist the Hume's 'heap' and 'bundle' metaphors. With this in mind, it is worth noting some complications in Hume's remarks on personal identity—the 'heap' and 'bundle' metaphors are not his final word on the issue.

While Hume's remarks on personal identity begin with the skeptical 'heap' and 'bundle' metaphors, his discussion in the main text of the *Treatise* concludes with an apparent solution, and a different metaphor. After considering classical puzzles about the identity of physical objects over time (1.4.6.8ff.), Hume announces that he intends to employ 'the same method of reasoning [...] which has so successfully explain'd the identity of plants, and animals, and ships, and houses' (1.4.6.15).³⁵ That method of reasoning is teleological: the identity of a ship or a plant is understood with reference to its *end*—a function that is sustained despite persistent change in the parts. Hume does not do us the favor of naming the end that is to inform the account of personal identity; given the broader context of the discussion, however, that end must be *cognition*.³⁶ The mind is to be considered as a cognitive capacity. The fluctuating parts are 'perceptions' or basic mental contents. Hume claims that they are 'link'd together by the relation of cause and effect, and mutually produce, destroy, influence, and modify each other' (1.4.6.19).

Putting the issue thus, Hume employs the quasi-mechanistic talk that is of a piece, surely, with his aspiration to be a 'Newton of the mind' (2000: 11): perceptions are like so many little balls bumping into one another—sometimes destroying one another, sometimes weakening or strengthening another's force or 'vivacity', sometimes combining to produce conglomerates of various kinds. This mechanistic talk is soon thrown into a new register, though, as Hume likens the mind to a *social* system:

I cannot compare the soul more properly to anything than to a republic or commonwealth, in which the several members are united by the

1 reciprocal ties of government and subordination, and give rise to other
 2 persons, who propagate the same republic in the incessant changes of its
 3 parts. And as the same individual may not only change its members, but
 4 also its laws and constitutions; in like manner the same person may vary
 5 his character and disposition, as well as his impressions and ideas,
 6 without losing his identity. (1.4.6.19)
 7

8 The coherence of representations, and hence the identity of the mind, is produced
 9 through something like governance. Hence there is a cognitive constitution that
 10 may be modified over time, and which may be continually assessed in light of its
 11 ability to preserve the coherence and harmony of the whole. Hume's proposal is
 12 to treat personal identity and this cognitive constitution as flip sides of a single
 13 coin.³⁷ And Kant, it seems, endorses this general strategy—though he goes
 14 further, in claiming that the potential coherence of the cognitive subject's
 15 representations rests upon a synthesis that is in some sense a priori.³⁸ This is the
 16 decisive issue, which I will address in the final section.

17 By dwelling upon some of the ambiguities of Hume's account, I mean to
 18 suggest that Hume arguably accepts the first two claims contained in Kant's
 19 apperception principle: that thinking is to be conceived as an activity of synthesis,
 20 and the coherence or unity of a subject's representations arises through the
 21 synthetic activity of thought. Moreover, I want to suggest that Hume accepts these
 22 claims on enlightenment grounds. Recall the ideal of enlightenment, as it is
 23 expressed through the three maxims considered above: namely, that a subject be
 24 able to recognize herself as the *source* of her own representations. Now, on the face
 25 of it, Hume might seem to ignore the enlightenment ideal almost entirely. After all,
 26 Hume takes *custom* to be 'the foundation of all of our judgments' (1.3.13.9).
 27 Customs are formed quite automatically: the regularities of nature, or of our social
 28 milieu, etch grooves or tracks along which the mind moves. At the stimulus of a
 29 given sense impression the mind moves with some degree of ease—depending on
 30 how well-worn the track is—to the corresponding idea (e.g. the anticipated future
 31 effect or the inferred past cause). So, although mental content is due to an activity
 32 of synthesis on the part of the subject, that synthesis follows tracks that are merely
 33 *imposed* upon the subject. The coherence of the cognitive subject is wrought by the
 34 automatic forces of custom. In this sense, Kant could accuse Hume for not really
 35 living up to the enlightenment ideal, at the heart of which lies the point that we
 36 must be able to regard ourselves as the *source* of our own thought.

37 But Hume does acknowledge this issue in a later development, when he points
 38 to the possibility of a subject's reflectively assessing the customs that shape her
 39 mental activity. When the mind is pushed forward by its own forces to anticipate
 40 some future effect (say) that in fact never materializes, a custom derails. The
 41 occasion of such a derailment introduces a kind of 'violence' into the mind, to use
 42 Hume's own term (1.3.11.4; 1.3.11.11): such violence threatens the coherence of
 43 the web of belief. The occasion of a derailment, in a suitably reflective cognitive
 44 subject, brings the relevant custom into view—whereas otherwise one's customs
 45 generally hum along quite unnoticed. Much goes along for the ride in an

1 association based upon custom: the color and smell of the liquid, say, that brings
 2 about a state of pleasant inebriation. If someone is confronted for the first time
 3 with a liquid that is now pale yellow instead of deep red, it is as if sand were
 4 thrown into the groove of the existing association, the track along which the mind
 5 tends to run. The movement of the mind will be obstructed to some degree or
 6 another, so that the subject now conceives less readily and vividly the anticipated
 7 inebriation. The custom has been brought into view; and should the subject find
 8 that the yellow liquid lifts her spirits just as well as the red, she is in a position to
 9 recognize that the color of the liquid is a 'superfluous' rather than an 'essential'
 10 circumstance and refine the custom accordingly (1.3.13.9).

11 Thus, while the coherence of a cognitive subject's representations is wrought by
 12 the automatic forces of custom, it is potentially refined and corrected through
 13 reflective self-examination when appropriate. Returning now to Hume's com-
 14 monwealth metaphor, we can find in it now a fundamental ambiguity: for who, or
 15 what, is the governor? This is the same as to ask: on what does the synthetic unity
 16 of representation depend? By taking custom to be the foundation of all of our
 17 judgments, Hume seems prepared to suppose that this governor is nothing other
 18 than nature herself. But Hume's remarks regarding the reflective assessment of
 19 custom suggest a different reading. Although the regularities of nature impress
 20 themselves upon the subject, creating grooves or tracks along which the mind
 21 runs, still the subject can represent these customs as rules, and ought to refine and
 22 correct them as appropriate. The unity and coherence of the subject's representa-
 23 tions would then be attributed to something that the subject *does*. Viewed in this
 24 context, the governor of the commonwealth would seem to be none other than the
 25 subject, as a robust cognitive agent.

26 This conception of cognitive agency needs to be in play if we are to take
 27 seriously the enlightenment requirement that a subject be able to recognize
 28 herself as the source of her cognitions. Hume acknowledges this requirement
 29 when he includes reflective self-assessment in his account of human cognition—
 30 so that we do not merely suffer the accretions of custom, but can (and should)
 31 reflectively examine our customs when they derail. However, this is an
 32 afterthought in his account, and it is overwhelmed by currents that pull in an
 33 opposing direction. Hume concludes his account of causal reasoning by
 34 suggesting that it is 'nothing but a wonderful and unintelligible instinct in our
 35 souls' (1.3.16.9). If the cognitive subject acts from a kind of instinct to reestablish
 36 unity and coherence of her representations when a custom derails, then we have
 37 lost the robust conception of agency that Hume seemed to invoke: when instinct
 38 governs the behavior of a creature it does so without that creature's needing to
 39 have any conception of this governance, or representation of its rules.

8. A Priori Synthesis

40 For Kant, if we are to take seriously the enlightenment requirement that a subject
 41 be able to recognize herself as the source of her cognitions, then we must also

1 invoke the idea of the *spontaneity* of the mind. Hence Kant first glosses this
 2 spontaneity as the mind's capacity to 'produce representations itself' (A51/B75).
 3 The synthesis that is responsible for the coherence of a subject's system of
 4 representations—and thus for her enjoying contentful representations at all—is
 5 due to the subject herself. This is the point of calling the synthesis appealed to in
 6 the apperception principle 'a priori'.³⁹

7 To understand what Kant means by 'a priori synthesis', it might help to
 8 recognize how close Hume comes to the idea—at least inasmuch as he suggests
 9 that personal identity rests upon cognitive agency. If Hume is committed to this
 10 much, then he should also grant the following. When a subject acts to restore
 11 unity and coherence among her representations, she does so with a certain end in
 12 view: for this much is entailed by the idea of intentional agency. By the lights of
 13 the commonwealth metaphor, we suppose that end to be *cognition*. So, if the
 14 cognitive subject is acting purposefully—refining her customs with this
 15 determinate end in view—she presumably has some general conception of the
 16 systematic order of representation that would constitute knowledge of nature,
 17 since that is what she is working towards.

18 That our cognitive activity should depend upon an idea of the whole of
 19 cognition is effectively what Kant maintains through the apperception principle.
 20 Kant is not asking us to imagine a priori synthesis as some kind of mysterious
 21 process; for he is not, he insists, trying to provide a genetic account of human
 22 cognition in the Transcendental Deduction (A86-7/B118-9). He is instead arguing
 23 that we must have a fundamental grasp of the systematic order of representation
 24 in order to be knowers at all. This fundamental grasp of the systematic order of
 25 representation is, in effect, the a priori synthesis of the apperception principle.
 26 Since synthesis is a combination of representations, an a priori synthesis would
 27 be a combination of representations that does not depend upon experience.
 28 Moreover, if we appeal to the synthesis model of thinking in order to account for
 29 mental content, then an a priori synthesis would presumably be a combination of
 30 representations that is the source of mental content as such. Finally, if we have in
 31 view some idea of a combination of representations that is the source of mental
 32 content as such, then this synthesis is a framework of the whole of cognition—not
 33 this or that concrete piece of it, but a systematic order in virtue of which this or
 34 that representation has the content that it has and can contribute to knowledge of
 35 material nature.

36 In the aftermath of the Deduction, this systematic order is articulated as a
 37 battery of 'principles of pure understanding' which concern the totality of nature
 38 as a law-governed whole.⁴⁰ Our extrapolation from Hume's commonwealth
 39 metaphor suggests that some such grasp of the systematic whole of representa-
 40 tion is required in order for there to be robust cognitive agency, which is itself
 41 integral to the enlightenment ideal Kant wants to defend. For Kant, this robust
 42 cognitive agency—the capacity of the subject to recognize herself as the source of
 43 her cognitions—is broadly attributed to the spontaneity of the mind.⁴¹

44 I have suggested that the apperception principle is dependent upon Kant's
 45 appeal to the spontaneity of the mind—so is everything about Kant's project

in the *Critique*, including the initial conception of reason, and the aspiration to vindicate a priori knowledge. To explain the precise nature of Kant's appeal to spontaneity is a further project, and one that I cannot take up here. Suffice it to say that Kant's appeal to spontaneity belongs to the broadly rationalist tradition, inasmuch as it invokes the idea of the mind's freedom from being determined by the causal order of material nature. So it is quite clear that Hume would not have set down the 'thorny path' of critical philosophy (*P* 4:367)—at least as Kant conceived it. It is all the more striking, then, that Kant engages at length with Hume as he seeks to establish the apperception principle in the Deduction. Yet if my account is correct, it makes sense. For Kant finds in Hume a philosopher who had some recognition—perhaps only dim—that the attention model of thinking cannot sustain the broad aspirations of the enlightenment project. So Kant engages with Hume to acknowledge this profound point of agreement, and then to show where they must part company, and why.⁴²

Melissa McBay Merritt
School of History and Philosophy
University of New South Wales
m.merritt@unsw.edu.au

Alison, H.

NOTES

¹ I have quoted here its complete name; see also the formulation in the title of §17 ('principle of the synthetic unity of apperception', B136), as well as the preliminary formulation that appears in §16 ('principle of the necessary unity of apperception', B135). The latter is incomplete because it omits the crucial point that the unity is *synthetic*, the significance of which I aim to address in the course of this paper.

² See §16 (B131–33); a more precise understanding of the principle will emerge in the course of this paper.

³ The two commentators who have most directly considered whether Kant provides, or has the resources to provide, an argument for the apperception principle are Paul Guyer and Henry Allison. Their differing accounts take the form of a debate about whether the apperception principle is 'analytic' or 'synthetic'. Allison argues that the apperception principle is analytic on the concept of a discursive intellect (see 1986; 1993; and 2004: 165–167). Guyer argues that the apperception principle would need to be synthetic a priori in order for it serve the purpose that Kant sets for it—unfortunately, however, Kant fails to provide an adequate argument for the principle (see Guyer 1980; and 1987: 131–154). It is not my intention to adjudicate their debate, although I will touch on aspects of it below.

⁴ Note, for example, James Van Cleve, who points to the 'principle of the unity of apperception' as the 'starting point of the Transcendental Deduction' (1999: 78); however, in the analysis that follows, he conflates this principle with the opening line of §16 (and omits the crucial qualifier '*synthetic*' that figures in Kant's complete formulation of the apperception principle). For a more nuanced version of the same oversight, consider Henry Allison, who complains about Patricia Kitcher's view that 'the doctrine of

apperception' is the conclusion of Kant's argument in the Deduction, when it is instead 'the initial premise of the Deduction' (1993: 237–8). But what is the 'doctrine of apperception'? Is it the apperception principle itself (which makes a claim about a priori synthesis), or some more general claim about apperception or self-consciousness? Although Allison evidently appreciates the difference (1993: 244), he does not consistently acknowledge it: in other work (2004: 163–4), he blurs the two.

⁵ This remark holds of the second-edition Deduction, which I shall focus on in this paper. Kant claims that the two versions of the Deduction differ only in the manner of exposition, and not in content (Bxxxviii). I believe this is correct, but it would distract from the line of thought that I wish to pursue here to take account of the significant differences in strategy and structure between the two versions. (Accounting for these differences would also be required to address the dispute between Guyer and Allison noted above.)

⁶ On the *Critique* as a project of self-knowledge, see Axi-xii, Bxxxv, A849/B877. The operative conception of reason is drawn from reason's claim to 'determine its object wholly a priori' (Bx). The *Critique of Pure Reason* sets out to vindicate this claim with respect to the *theoretical* employment of reason: that is, in the knowledge of material nature. The *Critique of Practical Reason* readdresses the claim of reason with regard to the *practical* employment of reason: that is, in the determination of the good to be brought about by action.

⁷ The three 'principal sources of prejudice' are not mutually exclusive (*JL* 9:76). For example, inclination will often encourage imitation, as Kant suggests when he points to the 'prejudices of prestige', where imitation is reinforced by our 'desire to imitate what is described to us as *great*' (9:78).

⁸ In 'What is Enlightenment?', Kant contrasts the disposition to think for oneself with a disposition to think in a way that would seem to express only the passive and fixed dispositions of a machine. Enlightenment, Kant concludes, makes one 'now more than a machine' (8:42; see also *JL* 9:76 for the same rhetorical emphasis). I examine this distinction between mechanism and spontaneity elsewhere (2009).

⁹ By 'common', Kant does not refer to what is ordinary: indeed, it may well be that these maxims are only rarely acknowledged in our *actual* cognitive practices. The three maxims specify the conditions of something like cognitive *health*, as Kant explains (5:293).

¹⁰ Kant presents the three maxims in *JL* (9:57), *KU* (5:294), and *Anth* (7:228 and 200). Although the presentation of the three maxims is largely consistent across these texts, Kant sometimes formulates the second maxim as 'to think in the position of another', rather than 'everyone else' (see 9:57 and 7:200). In 'What is Orientation in Thinking?', Kant glosses enlightenment as 'the maxim of always thinking for oneself' (8:146n). Kant does not explicitly refer to the maxims in 'What is Enlightenment?'; there the normativity of the notion of enlightenment is indicated by the juridical language employed throughout the essay. For other recent accounts of the three maxims, see O'Neill (1989: 46–47) and Deligiorgi (2002: 148–151).

¹¹ This is the opening line of §16. The first section of the Deduction—§15—serves an introductory function: it recalls a central thesis of the *Critique*, namely that sensibility and understanding are distinct, or heterogeneous, sources of cognition. This thesis is there recast in terms of a distinction between receptivity and spontaneity: 'The manifold of representations can be given in an intuition which is *merely* sensible, i.e. is nothing but receptivity, and the form of this intuition can lie a priori in our faculty of representation *without being anything other than the mode in which the subject is affected*' (B129; emphasis

added). What cannot be given through the senses is the 'combination of a manifold', for this is an 'act of spontaneity' and an 'action of the understanding' (B130). With this, Kant indicates that the overarching task of the Deduction is to account—in the face of this heterogeneity—for the necessary cooperation of sensibility and understanding in the production of knowledge. I argue for this interpretation of the task of the Deduction elsewhere (2006); in the present paper, my focus is the apperception principle, not the broader trajectory of the Deduction.

¹² I use this label for convenience, not to link the proposition to Descartes. (If anything, Kant means to underscore his distance from Cartesian philosophy of mind by putting emphasis on the possibility operator: it is not that the 'I think' *must* accompany all of my representations, but rather that it must be *able* to do so.)

¹³ I am using 'intellect' as a generic term, to refer to a capacity for thought. I will address Kant's terminology in greater detail in §5.

¹⁴ Guyer denies that the cogito statement is a self-evident proposition, pointing to hypnotic trances and talking during sleep as examples of representations that are evidently 'had', without being able to be 'had' in a self-conscious way (1987: 141). I take it that he means to point to certain kinds of mental states that, by nature, cannot be self-ascribed: mental goings-on in a hypnotic trance cannot be drawn into the fold of one's self-conscious representing. However, if I am right to suppose that the use of the first-person is meant to remind us that the *Critique* is a project of self-knowledge, then it follows that the epistemological concerns of the *Critique* would introduce a limitation on the scope of the cogito statement, so that it would pertain to representations inasmuch as they are potentially cognitively significant. It is not meant to apply to unconscious representations in trances or dreams.

¹⁵ Allison makes what may be a similar point when he notes that (what I refer to as) the cogito statement would hold even of an 'intuiting' or divine intellect; it is only on the presupposition of a discursive intellect that we get a claim about synthesis (1993: 244; see also 2004: 166). While this may be correct, it falls well short of providing an argument for the crucial claim about synthesis; Allison does not even provide an explanation of the entailment in question. We are left wondering why Kant should appeal to thinking as an activity of synthesis at all—a question I aim to address in this paper.

¹⁶ There is no *prima facie* reason to suppose that a certain mode of mental activity is fundamental to the mind: for there will of course be a variety of things that a mind does when it is active, and it may well be the case that not all of its activities are modes of a generic activity of thinking. However, the tendency to suppose that there must be a fundamental conception of the activity of the mind was widespread in the early modern era, perhaps due to the lingering influence of the Cartesian view that a mind is a substance, and thinking its essence or principal attribute.

¹⁷ As Descartes explains in the Third Replies, 'whatever is immediately perceived by the mind' counts as an 'idea' (1985: 127).

¹⁸ For the claim that a waking mind always thinking, see Locke, *Essay Concerning Human Understanding* II.xix.3 (1975: 227–228) and II.1.9 (108); on thinking as an activity of perception, see II.xxi.5 (236).

¹⁹ See *Essay* II.i.25 (1975: 118) and II.xix.1 (226). In these passages, Locke carefully maintains that the object of the mind's attention can only be to something *mental*, i.e. an 'idea'. However, in other passages, Locke speaks as if the first object of the mind's activity of attention were a *physical impression*—an impulse on the sensory organ, or a ruffling of the animal spirits in the brain, perhaps—seeming to suggest that this attention would itself yields mental contents: for such a passage, consider *Essay* II.ix.4 (144). I take it that part of

what Locke is struggling with here is the metaphysics of mind-body interaction, though I cannot take up the issue here.

²⁰ (I thank Joe Camp for putting this question to me, years ago.) Descartes also supposes that there must be some such correlation between sensory ideas and their physical causes—a point that rests upon the clear and distinct perception that God exists and is ‘not a deceiver’ (1985: 55). However, Descartes also recognizes another source of mental content, namely innate ideas. These, too, are ‘in’ a particular mind prior to that mind’s attention to them—they are ‘in’ the mind as the divine stamp upon created substance. As Descartes remarks in a letter of August 1641 to Hyperaspistes, the claim that thinking is the essence of mind does not entail that ‘the mind of an infant meditates on metaphysics in it its mother’s womb’: any particular subject, it seems, could remain forever like that infant, attending only to the ideas arising from the body (1991: 189–190). Ultimately, for Descartes, objective content rests upon the clear and distinct perception of innate ideas.

²¹ As Kant explains, this can be demonstrated by showing the categories to be conditions of the possibility of experience, thus avoiding the skeptical scenario in which appearances might ‘be so constituted that the understanding would not find them in accord with the conditions of its unity’ (A90/B123).

²² There is some controversy about what, exactly, ‘objective validity’ means for Kant. While most commentators take it to refer to objective purport, Kant sometimes seems to equate objective validity with *truth* (see, e.g. A788/B816). For present purposes, however, I shall set aside this difficulty: for at the very least objective validity means objective purport, which is all that is at issue for me here.

²³ Leibniz may see the problem with the attention model of thinking, at least in the Cartesian version. In the *New Essays*, Leibniz complains that the Cartesian view has the divine creator figuring as the source of necessary truths rather than the mind itself: ‘If all [the mind] had was the mere capacity to receive these items of knowledge—a passive power to do so, as indeterminate as the power of wax to receive shapes or of an empty page to receive words—it would not be the source of necessary truths, as I have just shown that it is’ (1981: 79). Leibniz’s account of necessary truths broadly concerns fundamental logical principles that are as ‘necessary for thought as muscles and tendons are for walking’: even though we may have no explicit recognition of them, nevertheless we tacitly endorse them inasmuch as we exercise our reason at all (83–4). I take this to have been influential in Kant’s decision to take pure general logic as a starting point for the ‘transcendental logic’ that he begins to work out in the Transcendental Deduction—but this is a broader issue that I cannot enter into here.

²⁴ Kant reminds us of the distinction in the Deduction (§22, B146). On the meaning of objective validity, see note 22 above.

²⁵ Thus Kant glosses the ‘understanding in general’ as ‘the faculty of rules’ (A132/B171).

²⁶ How are we to understand the bare notion of *representation* in Kant’s usage? It is tempting to suppose that representation is the object of thought, or what figures in consciousness. But the notion of representation is something more rudimentary than these glosses would allow: for only a creature with a faculty of understanding *thinks*, on Kant’s view; and not every representation is had with consciousness. Note that in Kant’s ‘progression’ or ‘*Stufenleiter*’ of terms concerned with representation in the *Critique*, the term *representation* is accepted as a genus and only its species are distinguished (A320/B376–7).

²⁷ See, e.g., Leibniz (1981: 53–4). That Kant might have Leibnizian *petites perceptions* in mind is suggested in *Anth* §5 (7:135).

²⁸ See *Anth* (7:117). (Kant's gloss on the notion of consciousness also invites a regress problem, though I will not take up that issue here.)

²⁹ If we were to give verbal expression to the second and third grades in Kant's list, we might resort to: *here something*; and *this is not that*.

³⁰ Comprehension is scientific cognition. In the passage under consideration, Kant points to mathematics—widely accepted as scientific cognition *par excellence*, and which Kant takes to be a science of nature (see *Critique* §22, B147).

³¹ Kant sometimes refers to the 'higher cognitive faculty' as the 'understanding in general'; he takes it to be comprised of understanding, the power of judgment (i.e. to subsume representations under rules), and reason, the faculty of inference (A130-1/B169). This division corresponds to the then-standard division in logic textbooks—addressing concepts, judgment, and inference in that order.

³² I cannot say 'of thinking' because thought, for Kant, can only belong to a creature capable of cognition.

³³ Hume's account of personal identity is not straightforward: his initial skepticism about personal identity is followed by an apparent solution the problem of personal identity—which solution is then later dismissed in the Appendix. I will discuss these developments in §7: for I take the non-skeptical phase of Hume's account of personal identity to indicate that he comes closer to the apperception principle than is generally appreciated. That said, however, it is not my aim in this paper to provide a complete account of Hume's views on personal identity.

³⁴ The passages about personal identity are found in *Treatise* 1.4.6 (2007: 164–171). Although this is not a text that Kant would have read directly, as Robert Paul Wolff points out, he would have known of Hume's provocative suggestion that the self is nothing more than a 'heap' or a 'bundle' of perceptions through the 1772 German translation of James Beattie's *Essay on the Nature and Immutability of the Truth*, which quotes in full the passages in which the famous 'heap' and 'bundle' metaphors occur (see Wolff 1960: 117 and 120–1). Kant would also have been aware of the allusion to Hume's 'heap' metaphor in Johann Nicholas Tetens' *Philosophische Versuche*—which Tetens rejects: 'Mein Ich ist ein Eins, nicht ein Haufen von mehreren Dingen' (1777: 178).

³⁵ This promise of an explanation is not meant to mitigate Hume's view that 'The identity, which we ascribe to the mind of man, is only a fictitious one' (2007: 169). The account of personal identity follows upon Hume's account of the similar 'fiction' regarding the 'continued existence' of material bodies in *Treatise* 1.4.2.

³⁶ I am using 'cognition' in a somewhat loose sense here. Hume often speaks as if only reasoning concerning 'relations of ideas'—i.e. what we would refer to as analytic judging, yielding claims that hold of necessity—should be admitted as knowledge; reasoning concerning 'matters of fact' does not admit of certainty, rests only on probabilities, and yields only varying degrees of *belief* (for an example of this way of talking, see, e.g. 1.3.13.19). When I say here that the relevant end must be 'cognition', I am using the term in a broader sense that would include probabilistic reasoning concerning matters of fact.

³⁷ In the *Treatise's* Appendix, Hume disparages the idea that the commonwealth metaphor might provide some kind of solution to the problem of personal identity. The commonwealth metaphor of course relies upon the idea that a subject has a multiplicity of representations. In the Appendix, Hume has us consider the mind as 'reduc'd even below the life of an oyster', having 'only one perception, as of thirst or hunger' (2007: 399)—not even fluctuations between the two. Hume suggests that since it would be (intuitively) absurd to attribute selfhood to such a mind, 'the addition of other perceptions' to this subject 'could never give you that notion', either. Obviously, if the mind has only one

continual perception, there is nothing to compare, connect, or associate: there is no synthetic activity, and the attention model of thinking is forced upon us.

³⁸ Edwin McCann (1985: 75) notes that Kant draws upon a 'Humean' view of personal identity—one, at least, that 'renders the identity of the self in terms of relations between its representations'. However, even if we should restrict our attention to Hume's commonwealth metaphor, it would not follow that he has quite the same conception of personal identity as Kant. The fact that according to Hume the activity of synthesis follows the automatic path of custom entails that the analogy with governance is somewhat strained, given that the latter suggests autonomy or self-legislation. This, indeed, is another way of getting at the fundamental break between Hume and Kant.

³⁹ Guyer glosses the a priori synthesis invoked in the apperception principle as the activity of mind responsible for 'an actual *imposition* of order on nature' (1980: 207 and *passim*). On my reading, the apperception principle concerns the conditions of having contentful representations at all; it does not say anything about the imposition of order on nature. The apperception principle is systematically prior to any claims about nature, which arise from an *application* of the apperception principle to particular facts about human sensibility. This move is made in the second half of the Deduction, yielding a further principle: namely that whatever may be represented spatio-temporally is necessarily subject to categorial thought. This result provides the basis for Kant's subsequent enumeration of a battery of 'principles of pure understanding' in the aftermath of the Deduction, which themselves articulate a general conception of nature as a law-governed whole. As the apperception principle is systematically prior to these developments, it is itself silent on the topic of nature.

⁴⁰ Kant first alludes to this battery of principles, and hence a priori synthesis, when he points to the enlightened practices of experimental natural science in the B Preface: 'Reason, holding in one hand its principles, according to which alone agreement among appearances can be admitted as laws, and in the other hand the experiment which it has thought out in accordance with these principles, must approach nature in order to be taught by it. But it must not do so in the character of a pupil, who lets whatever the teacher wants to say be recited to him, but rather like an appointed judge who compels the witnesses to answer the questions which he puts to them' (Bxiv).

⁴¹ Susan Hurley charges that Kant falls prey to 'the myth of the giving': Kant 'fails to come to grip with the fact that intentional acts themselves, however spontaneous they may be, have content and thus presuppose the unity of the concepts or the object that provide their content just as much as perceptual experience does' (1994: 154). On her view, Kant appeals to an active synthesis to explain the unity of the subject, 'as if unity could not be a feature of the data of sensibility on their own, and, more importantly still, as if agency did not already depend upon unity' (163). Hurley's charge depends upon our supposing that a priori synthesis is intentional action; however, it is not obvious that we should conceive of a priori synthesis in this way—and if not, then no question arises as to its content. Elsewhere (2009), I argue that Kant's appeal to spontaneity must be understood in terms of the normative framework of our cognitive lives—an approach that Hurley dismisses rather quickly (see 1994: 152). According to my account here, Kant's account of mental content emerges when we appreciate the apperception principle *in light of* that broader normative context. It is not meant to be established through the appeal to a priori synthesis directly.

⁴² I am grateful to many individuals—both known and anonymous—who have commented on this paper, or one of its ancestors. Special thanks are due to Markos Valaris, who helped me see it through its several incarnations.

REFERENCES

References to Kant's works, with the exception of the *Critique of Pure Reason*, refer to the volume and page of the German Academy of Sciences edition. References to the *Critique of Pure Reason* follow the pagination of the first and second editions, abbreviated A/B accordingly. Other abbreviations are as follows: *Anth* = *Anthropology from a Pragmatic Point of View*; *JL* = *Jäsche Logic*; *KU* = *Critique of Judgment*; *P* = *Prolegomena to Any Future Metaphysics*. Translations from the *Critique of Pure Reason* are my own; translations from other works are drawn from the Cambridge Edition of the Works of Immanuel Kant.

References to Hume's *Treatise* are given by book, part, section, and (where applicable) paragraph as indicated in the Clarendon Edition of the Works of David Hume.

Q4

Alison, H. (1986), 'Reflections on the B-Deduction', *Southern Journal of Philosophy*, 25(Supplement): 1–15.

— (1993), 'Apperception and Analyticity in the B-Deduction', *Grazer Philosophische Studien*, 44: 233–52.

— (2004), *Kant's Transcendental Idealism*, Revised and Enlarged Edition. New Haven and London: Yale University Press.

Deligiorgi, K. (2002), 'Universalisability, Publicity, and Communication: Kant's Conception of Reason', *European Journal of Philosophy*, 10: 143–59.

Descartes, R. (1985), *The Philosophical Writings of Descartes: Volume II*, trans. J. Cottingham, R. Stoothoff, D. Murdoch. Cambridge: Cambridge University Press.

— (1991), *The Philosophical Writings of Descartes: Volume III, The Correspondence*, trans. J. Cottingham, R. Stoothoff, D. Murdoch and A. Kenny.

Guyer, P. (1980), 'Kant on Apperception and A Priori Synthesis', *American Philosophical Quarterly*, 17: 205–12.

— (1987), *Kant and the Claims of Knowledge*. Cambridge: Cambridge University Press.

Q1

Hume, D. (2000), *An Enquiry concerning Human Understanding*, ed. T. Beauchamp. Oxford: Oxford University Press.

— (2007), *A Treatise of Human Nature: A Critical Edition*, ed. D. F. Norton and M. J. Norton. Volume 1: Texts. Oxford: Oxford University Press.

Hurley, S. (1994), 'Kant on Spontaneity and the Myth of the Giving', *Proceedings of the Aristotelian Society, New Series*, 94: 137–164.

Kant, I. (1902-), *Gesammelte Schriften*, ed. Königlich Preußische Akademie der Wissenschaften, later the Deutsche Akademie der Wissenschaften zu Berlin. 29 vols Berlin: Walter de Gruyter (and predecessors).

Leibniz, G. (1981), *New Essays on Human Understanding*, trans. and ed. P. Remnant and J. Bennett. Cambridge: Cambridge University Press.

Locke, J. (1975), *An Essay concerning Human Understanding*, ed. P. H. Nidditch. Oxford: Clarendon.

McCann, E. (1985), 'Skepticism and Kant's B Deduction', *History of Philosophy Quarterly*, 2: 71–89.

Q2

Merritt, M. (2006), 'Science and the Synthetic Method of the *Critique of Pure Reason*', *Review of Metaphysics*, 59: 517–39.

Q3

— (2009), 'Reflection, Enlightenment, and the Significance of Spontaneity in Kant', *British Journal for the History of Philosophy*, 17 (forthcoming).

O'Neill, O. (1989), *Constructions of Reason*. Cambridge: Cambridge University Press.

- 1 Tetens, J. (1777), *Philosophische Versuche über die Menschliche Natur und Ihre Entwicklung*,
2 Second volume. Leipzig: Weidmanns Erben und Reich.
3 Van Cleve, J. (1999), *Problems from Kant*. Oxford and New York: Oxford University Press.
4 Wolff, R. (1960), 'Kant's Debt to Hume via Beattie', *Journal of the History of Ideas*, 21:
5 117–23.
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33

